

# CROWDSOURCING FOR LINGUISTIC FIELD RESEARCH AND E-LEARNING



Dissertation  
an der Fakultät für Mathematik, Informatik und Statistik  
der Ludwig-Maximilians-Universität München

vorgelegt von  
Diplom-Informatiker  
Fabian Kneißl

München, den 25. Februar 2014

Erstgutachter: Prof. Dr. François Bry  
Zweitgutachter: PD Dr. Georg Groh  
Tag der mündlichen Prüfung: 1. April 2014

## Eidesstattliche Versicherung

(Siehe Promotionsordnung vom 12.07.11, § 8, Abs. 2 Pkt. .5.)

Hiermit erkläre ich an Eidesstatt, dass die Dissertation von mir selbstständig, ohne unerlaubte Beihilfe angefertigt ist.

**Kneißl, Fabian**

-----  
Name, Vorname

**München, den 25.2.2014**

-----  
Ort, Datum

-----  
Unterschrift Doktorand/in



# ABSTRACT

Crowdsourcing denotes the transfer of work commonly carried out by single humans to a large group of people. Nowadays, crowdsourcing is employed for many purposes, like people contributing their knowledge to Wikipedia, researchers predicting diseases from data on Twitter, or players solving protein folding problems in games. Still, there are areas for which the application of crowdsourcing has not yet been investigated thoroughly. This thesis examines crowdsourcing for two such areas: for empirical research in sciences oriented on humans –focusing on linguistic field research– and for e-learning.

Sciences oriented on humans –like linguistics, sociology, or art history– depend on empirical research. For example, in traditional linguistic field research researchers ask questions and fill in forms. Such methods are time-consuming, costly, and not free of biases. This thesis proposes the application of crowdsourcing techniques to overcome these disadvantages and to support empirical research in getting more efficient. Therefore, the concept of a generic market for trading with symbolic goods and speculating on their characteristics in a playful manner, called *Agora*, is introduced. *Agora* aims to be an “operating system” for social media applications gathering data. Furthermore, the Web-based crowdsourcing platform *metropolitalia* has been established for hosting two social media applications based upon *Agora*: *Mercato Linguistico* and *Poker Parole*. These applications have been conceived as part of this thesis for gathering complementary data and meta-data on Italian language varieties. *Mercato Linguistico* incites players to express their own knowledge or beliefs, *Poker Parole* incites players to make conjectures on the contributions of others. Thereby the primary meta-data collected with *Mercato Linguistico* are enriched with secondary, reflexive meta-data from *Poker Parole*, which are needed for studies on the perception of languages. An evaluation of the data gathered on *metropolitalia* exhibits the viability of the market-based approach of *Agora* and highlights its strengths.

E-learning is concerned with the use of digital technology for learning, nowadays especially via the Internet. This thesis investigates how e-learning applications can support students with association-based learning and lecturers with teaching. For that, a game-like e-learning tool named *Termina* is proposed in this thesis. From the data collected with *Termina* association maps are constructed. An association map is a simplified version of a concept map, in which concepts are represented as rectangles and relationships between concepts as links. They constitute an abstract comprehension of a topic. Students profit from the association maps’ availability, learn from other participating students, and can track their own learning progress. Lecturers gain insights into the knowledge and into potential misunderstandings of their students. An evaluation of *Termina* and the collected data along a university course exhibits *Termina*’s usefulness for both students and lecturers.

The main contributions of this thesis are (1) a literature review over collective intelligence, crowdsourcing, and related fields, (2) a model of a generic market for gathering data for empirical research efficiently, (3) two applications based on this model and results of an evaluation of the data gathered with them, (4) the game-like e-learning tool Termina together with insights from its evaluation, and (5) a generic software architecture for all aforementioned applications.

## ZUSAMMENFASSUNG

Crowdsourcing bezeichnet die Auslagerung von Arbeit an eine Gruppe von Menschen zur Lösung eines Problems. Heutzutage wird Crowdsourcing für viele Zwecke verwendet, zum Beispiel tragen Leute ihr Wissen zu Wikipedia bei, Wissenschaftler sagen Krankheiten anhand von Twitter-Daten vorher oder Spieler lösen Proteinfaltungsprobleme in Spielen. Es gibt dennoch Gebiete, für die der Einsatz von Crowdsourcing noch nicht gründlich untersucht wurde. Diese Arbeit untersucht Crowdsourcing für zwei solche Gebiete: für empirische Forschung in auf den Menschen bezogenen Wissenschaften mit Fokus auf linguistischer Feldforschung sowie für E-Learning.

Auf den Menschen bezogene Wissenschaften wie Linguistik, Soziologie oder Kunstgeschichte beruhen auf empirischer Forschung. In traditioneller linguistischer Feldforschung zum Beispiel stellen Wissenschaftler Fragen und füllen Fragebögen aus. Solche Methoden sind zeitaufwändig, teuer und nicht unbefangen. Diese Arbeit schlägt vor, Crowdsourcing-Techniken anzuwenden, um diese Nachteile zu überwinden und um empirische Forschung effizienter zu gestalten. Dazu wird das Konzept eines generischen Marktes namens *Agora* für den Handel mit symbolischen Gütern und für die Spekulation über deren Charakteristika eingeführt. Agora ist ein generisches “Betriebssystem” für Social Media Anwendungen. Außerdem wurde die Internet-basierte Crowdsourcing-Plattform *metropolititalia* eingerichtet, um zwei dieser Social Media Anwendungen, die auf Agora basieren, bereitzustellen: *Mercato Linguistico* und *Poker Parole*. Diese Anwendungen wurden als Teil dieser Arbeit entwickelt, um komplementäre Daten und Metadaten über italienische Sprachvarietäten zu sammeln. *Mercato Linguistico* regt Spieler dazu an, ihr eigenes Wissen und ihre Überzeugungen auszudrücken. *Poker Parole* regt Spieler dazu an, Vermutungen über die Beiträge anderer Spieler anzustellen. Damit werden die mit *Mercato Linguistico* gesammelten primären Metadaten mit reflexiven sekundären Metadaten aus *Poker Parole*, die für Studien über die Wahrnehmung von Sprachen notwendig sind, bereichert. Eine Auswertung der auf *metropolititalia* gesammelten Daten zeigt die Zweckmäßigkeit des marktbasierten Ansatzes von Agora und unterstreicht dessen Stärken.

E-Learning befasst sich mit der Verwendung von digitalen Technologien für das Lernen, heutzutage vor allem über das Internet. Diese Arbeit untersucht, wie E-Learning-Anwendungen Studenten bei assoziationsbasiertem Lernen und

Dozenten bei der Lehre unterstützen können. Dafür wird eine Spiel-ähnliche Anwendung namens *Termina* in dieser Arbeit eingeführt. Mit den über *Termina* gesammelten Daten werden Association-Maps konstruiert. Eine Association-Map ist eine vereinfachte Variante einer Concept-Map, in der Begriffe als Rechtecke und Beziehungen zwischen Begriffen als Verbindungslinien dargestellt werden. Sie stellen eine abstrakte Zusammenfassung eines Themas dar. Studenten profitieren von der Verfügbarkeit der Association-Maps, lernen von anderen Studenten und können ihren eigenen Lernprozess verfolgen. Dozenten bekommen Einblicke in den Wissensstand und in eventuelle Missverständnisse ihrer Studenten. Eine Evaluation von *Termina* und der damit gesammelten Daten während eines Universitätskurses bestätigt, dass *Termina* sowohl für Studenten als auch für Dozenten hilfreich ist.

Die Kernbeiträge dieser Arbeit sind (1) eine Literaturrecherche über kollektive Intelligenz, Crowdsourcing und verwandte Gebiete, (2) ein Modell eines generischen Marktes zur effizienten Sammlung von Daten für empirische Forschung, (3) zwei auf diesem Modell basierende Anwendungen und Ergebnisse deren Evaluation, (4) die Spiel-ähnliche E-Learning-Anwendung *Termina* zusammen mit Einblicken aus dessen Evaluation und (5) eine generische Softwarearchitektur für alle vorgenannten Anwendungen.





# PUBLICATIONS

Some ideas and figures presented in this thesis previously appeared in the following publications:

- [1] François Bry, Fabian Kneißl, Thomas Krefeld, Stephan Lücke and Christoph Wieser. *A Crowdsourcing Platform for Italian Linguistic Field Research*. Research Report. Munich, Germany: Institute for Informatics, Ludwig-Maximilians-Universität München, 2013.
- [2] François Bry, Fabian Kneißl, Thomas Krefeld, Stephan Lücke and Christoph Wieser. “Crowdsourcing for a Geographical and Social Mapping of Italian Dialects”. In: *Proceedings of the 2nd International Workshop on Social Media for Crowdsourcing and Human Computation (SOHUMAN)*. Paris, France, 2013.
- [3] François Bry, Fabian Kneißl and Christoph Wieser. “Field Research for Humanities with Social Media: Crowdsourcing and Algorithmic Data Analysis”. In: *Proceedings of 4. Workshop Digitale Soziale Netze im Rahmen der 41. Jahrestagung der Gesellschaft für Informatik (DSN)*. Berlin, Germany, 2011.
- [4] Fabian Kneißl and François Bry. *Borsa Parole - A Market for Linguistic Speculation*. Research Report. Munich, Germany: Institute for Informatics, Ludwig-Maximilians-Universität München, 2012.
- [5] Fabian Kneißl and François Bry. “Fostering Concept Maps Awareness as a Means to Learning”. In: *Proceedings of the Third International Conference on Social Computing and Its Applications (SCA)*. Karlsruhe, Germany, 2013.
- [6] Fabian Kneißl and François Bry. “MetropolItalia: A Crowdsourcing Platform for Linguistic Field Research”. In: *Proceedings of the IADIS International Conference WWW/Internet (ICWI)*. Madrid, Spain, 2012.
- [7] Fabian Kneißl and François Bry. “Speculation and Bluff for Linguistic Field Research - Social Software and Media for Data Gathering”. In: *Proceedings of 5. Workshop Digitale Soziale Netze im Rahmen der 42. Jahrestagung der Gesellschaft für Informatik (DSN)*. Braunschweig, Germany, 2012.
- [8] Stephan Leutenmayr, Fabian Kneißl, Sven Ziemer and François Bry. “Gameful Markets for Collaboration and Learning”. In: *Proceedings of Disco, Workshop on Human Computation and Machine Learning in Games at HCOMP*. Palm Springs, CA, USA, 2013.

- [9] Bartholomäus Steinmayr, Christoph Wieser, Fabian Kneißl and François Bry. “Karido: A GWAP for Telling Artworks Apart”. In: *Proceedings of the 16th International Conference on Computer Games (CGAMES)*. Louisville, KY, USA, 2011.

# ACKNOWLEDGEMENTS

This thesis would not have been possible without the enduring support of many people. First, I would like to thank my supervisor François Bry for introducing me to the topic of crowdsourcing and for the countless discussions and his continuous motivation. I also thank Georg Groh for his time as an external reviewer and his comprehensive comments.

The members of my research unit have been very supportive in the last three years. In particular, I would like to thank Norbert Eisinger for consulting me during my thesis and for proofreading parts of my thesis. And also, I would like to thank Steffen Hausmann, Martin Josko, Alexander Pohl, Ingeborg von Troschke, Christoph Wieser and the other members of faculty for making the office days enjoyable during the last years and for many interesting discussions.

Furthermore, my thanks go to Hubertus Kohle, Thomas Krefeld, Stephan Lücke, Klaus Schulz and all other project members of the Play4Science project, who have been very supportive and who gave me feedback and shared their ideas. I also thank Caterina Campanella, Laura Commare, Silvia Cramerotti, Debora Francione, Katharina Jakob, and Alessandra Puglisi for their support of the metropolitalia platform. My thanks also go to the German Research Foundation (DFG) for funding me as part of the Play4Science project number 578416, which allowed me to pursue research, attend conferences and get in touch with fellow researchers.

Several aspects of this work have been done together with students whose theses I supervised: Niels Becker, Romy Buchschmid, Katharina Krug, André Reichstaller.

Last but not least, I would like to thank my fiancée Anja Abel for her patience, her understanding, and her ongoing support, Franz Kneißl for proofreading parts of this thesis and the rest of my family for encouraging me.



# CONTENTS

<b>I</b>	<b>INTRODUCTION AND RELATED WORK</b>	<b>1</b>
1	INTRODUCTION	3
1.1	Contributions . . . . .	4
1.2	Structure of this Thesis . . . . .	5
2	RELATED WORK	7
2.1	Collective Intelligence . . . . .	7
2.1.1	Social Computing . . . . .	8
2.1.2	Crowdsourcing . . . . .	10
2.1.3	Human Computation . . . . .	12
2.1.4	Games With a Purpose (GWAPs) . . . . .	13
2.1.5	Prediction Markets . . . . .	15
2.2	Collective Intelligence Applications . . . . .	16
2.2.1	Linguistic Applications . . . . .	16
2.2.2	Market-Based Applications . . . . .	17
2.2.3	Applications for Learning . . . . .	18
2.3	Gaming Techniques in Standalone and Collaborative Applications	19
2.3.1	Serious Games . . . . .	19
2.3.2	Gamification . . . . .	20
2.4	E-Learning . . . . .	20
2.4.1	General Concepts of E-Learning . . . . .	20
2.4.2	Learning Analytics . . . . .	21
2.4.3	Concept Maps . . . . .	21
<b>II</b>	<b>LINGUISTIC FIELD RESEARCH</b>	<b>23</b>
3	METROPOLITALIA: A CROWDSOURCING PLATFORM FOR ITALIAN LINGUISTIC FIELD RESEARCH	25
3.1	Metropolitalia's Goals . . . . .	26
3.2	Linguistic Field Research on the Italian Language . . . . .	26
3.2.1	Traditional Linguistic Field Research . . . . .	26
3.2.2	History of the Italian Language . . . . .	27
3.2.3	Humans' Interest in Language . . . . .	28
3.3	Focus on Italian Language Varieties . . . . .	29
4	AGORA: A MARKET-BASED OPERATING SYSTEM FOR CROWDSOURCING APPLICATIONS	31
4.1	Core Concepts . . . . .	32
4.1.1	Symbolic Goods under Consideration . . . . .	34
4.1.2	Characteristics of Symbolic Goods . . . . .	35
4.1.3	Characterisations, Assessments, and Further Actions . . . . .	37
4.2	The Role of Play-Money in Agora . . . . .	40
4.2.1	Monetary Value of Assessments . . . . .	40

4.2.2	Dividends . . . . .	43
4.2.3	Trading with Assessments . . . . .	43
4.3	User Interactions . . . . .	44
4.3.1	Creating a Symbolic Good . . . . .	45
4.3.2	Characterising a Symbolic Good . . . . .	46
4.3.3	Trading . . . . .	47
4.3.4	Searching and Browsing . . . . .	48
4.4	Agora's Similarity to a Financial Market . . . . .	49
4.5	Ensuring the Quality of the Gathered Data . . . . .	50
4.5.1	Strategies to Prevent Cheating . . . . .	51
4.5.2	Gold Standard Assessments . . . . .	51
4.5.3	Filtering Data . . . . .	52
5	MERCATO LINGUISTICO AND POKER PAROLE – TWO COMPLEMEN- TARY CROWDSOURCING APPLICATIONS BUILT WITH AGORA	55
5.1	Mercato Linguistico . . . . .	55
5.1.1	User Interactions . . . . .	56
5.1.2	Customisation of Agora's Generic Aspects . . . . .	59
5.1.3	Success on Mercato Linguistico: Keynes' Beauty Contest	64
5.2	Poker Parole . . . . .	65
5.2.1	User Interactions . . . . .	65
5.2.2	Customisation of Agora's Generic Aspects . . . . .	66
5.2.3	Success on Poker Parole: Speculation with Rarely Known Phrases . . . . .	69
5.3	User Incentives . . . . .	69
5.4	The Cold Start Problem . . . . .	71
5.5	Classification in the Field of Collective Intelligence . . . . .	72
5.6	Potential of Agora in Art History . . . . .	73
6	EVALUATION OF DATA GATHERED WITH MERCATO LINGUISTICO AND POKER PAROLE	75
6.1	Temporal Development . . . . .	76
6.2	Quantitative Breakdown . . . . .	79
6.2.1	Total Amount of Gathered Data . . . . .	79
6.2.2	Gathered Data per User . . . . .	81
6.2.3	Gathered Characterisations per Phrase . . . . .	82
6.3	Evaluation of the Assessment's Estimated Agreement Proportion	83
6.4	Quality of Gathered Data for Linguistic Field Research . . . . .	86
III	E-LEARNING	89
7	THE GAME-LIKE E-LEARNING TOOL TERMINA	91
7.1	Gathering Semantic Associations . . . . .	92
7.1.1	Deploying Seed Data . . . . .	93
7.1.2	Classification of Associated Terms . . . . .	94
7.2	Termina's Modes of Operation . . . . .	95
7.2.1	Play Mode . . . . .	95
7.2.2	Practise Mode . . . . .	97
7.3	Benefits of Using Termina . . . . .	99

7.3.1	Association-Based Learning . . . . .	99
7.3.2	Feedback for Students . . . . .	100
7.3.3	Feedback for Lecturers . . . . .	101
7.4	Construction of Association Maps . . . . .	102
7.5	Termina for Expert Communities . . . . .	103
8	EVALUATION OF TERMINA ALONG A UNIVERSITY COURSE	105
8.1	Data Quantity . . . . .	105
8.2	Data Quality . . . . .	107
8.3	User Survey . . . . .	110
8.4	Conclusions . . . . .	111
IV	SOFTWARE ARCHITECTURE	113
9	SOFTWARE ARCHITECTURE	115
9.1	Seam Framework as Basis . . . . .	115
9.1.1	Foundations on Java EE 5 . . . . .	115
9.1.2	Extended Persistence Context . . . . .	116
9.1.3	Context and State Management . . . . .	116
9.1.4	Dependency Bijection . . . . .	117
9.1.5	Automated Testing . . . . .	117
9.1.6	Web 2.0 Applications with Ajax . . . . .	118
9.2	Application Server . . . . .	118
9.3	Modular Concept . . . . .	120
9.4	Database Structure . . . . .	120
9.5	Search with Solr . . . . .	123
V	CONCLUSION	125
10	CONCLUSION AND FUTURE WORK	127
A	APPENDIX	131
A.1	Evaluation of Mercato Linguistico and Poker Parole . . . . .	131
A.1.1	Gathered Data per User . . . . .	131
A.1.2	Characterisations per Phrase . . . . .	135
A.1.3	The Assessment's Agreement Proportion . . . . .	137
A.2	Evaluation of Termina . . . . .	138
A.2.1	Gathered Data . . . . .	138
A.2.2	User Survey . . . . .	140
	BIBLIOGRAPHY	145





# LIST OF FIGURES

Figure 1	Diagram of collective intelligence . . . . .	9
Figure 2	Concept map about concept maps . . . . .	22
Figure 3	Agora’s data schema as UML diagram . . . . .	33
Figure 4	Functions $value_{linear}$ and $value_{nd}$ . . . . .	42
Figure 5	Screenshot of Mercato Linguistico during the choice of a region	56
Figure 6	Screenshot of Mercato Linguistico for adding new phrases . .	57
Figure 7	Stepwise focusing on smaller regions . . . . .	58
Figure 8	Screenshot of Mercato Linguistico for reviewing own assess- ments . . . . .	59
Figure 9	Screenshot of Mercato Linguistico for purchasing assessments	60
Figure 10	Function $value_{nd}$ . . . . .	64
Figure 11	Function $5 \cdot \max(1 - 5 \cdot agreement_{pp}(a), 0)$ . . . . .	69
Figure 12	Diagram of collective intelligence with Mercato Linguistico and Poker Parole . . . . .	72
Figure 13	Active participants on metropolitalia per month . . . . .	76
Figure 14	New phrases added to metropolitalia per month . . . . .	77
Figure 15	Characterisations gathered with Mercato Linguistico per month	78
Figure 16	Assessments gathered with Mercato Linguistico and Poker Parole per month . . . . .	79
Figure 17	Actions users perform in Mercato Linguistico . . . . .	81
Figure 18	Distribution of geographical characterisations by users on Mercato Linguistico . . . . .	82
Figure 19	Distribution of geographical characterisations per phrase on Mercato Linguistico . . . . .	83
Figure 20	Histogram of the estimated agreement proportion of assess- ments in Mercato Linguistico . . . . .	84
Figure 21	Histogram of the difference between the estimated agreement proportion and the calculated agreement of assessments in Mercato Linguistico . . . . .	85
Figure 22	Histogram of the difference between the estimated agreement proportion and the calculated agreement of assessments in Mercato Linguistico for which the estimated agreement pro- portion is not 10 . . . . .	85
Figure 23	Screenshot of search results on metropolitalia . . . . .	87
Figure 24	Screenshot of a Termina session . . . . .	92
Figure 25	Screenshot of the difficulty settings in Termina . . . . .	96
Figure 26	Screenshot of the summary displayed at the end of a Termina session . . . . .	98
Figure 27	Screenshot of the start page of Termina . . . . .	100
Figure 28	Association map of a concept subset on markup languages . .	103
Figure 29	Active participants on Termina per week . . . . .	106

Figure 30	Association map of a concept subset on function composition	109
Figure 31	Results of the Termina survey . . . . .	111
Figure 32	Directory tree of the gwap project . . . . .	119
Figure 33	Files and directories of the gwap EAR . . . . .	121
Figure 34	Entity-relationship diagram of the database . . . . .	122
Figure 35	Distribution of phrases added by users on Mercato Linguistico and Poker Parole. . . . .	131
Figure 36	Distribution of geographical assessments by users on Mercato Linguistico . . . . .	132
Figure 37	Distribution of word selection characterisations by users on Mercato Linguistico . . . . .	132
Figure 38	Distribution of age characterisations by users on Mercato Linguistico . . . . .	133
Figure 39	Distribution of gender characterisations by users on Mercato Linguistico . . . . .	133
Figure 40	Distribution of level of education characterisations by users on Mercato Linguistico . . . . .	134
Figure 41	Distribution of assessments by users on Poker Parole . . . . .	134
Figure 42	Distribution of word selection characterisations per phrase on Mercato Linguistico . . . . .	135
Figure 43	Distribution of age characterisations per phrase on Mercato Linguistico . . . . .	135
Figure 44	Distribution of gender characterisations per phrase on Mercato Linguistico . . . . .	136
Figure 45	Distribution of level of education characterisations per phrase on Mercato Linguistico . . . . .	136
Figure 46	Distribution of estimated agreement proportions of assessments	137
Figure 47	Completed game rounds per week . . . . .	138
Figure 48	Stated associated terms per week . . . . .	138
Figure 49	Answers to the survey question number 1 . . . . .	140
Figure 50	Answers to the survey question number 2 . . . . .	140
Figure 51	Answers to the survey question number 3 . . . . .	141
Figure 52	Answers to the survey question number 4 . . . . .	141
Figure 53	Answers to the survey question number 5 . . . . .	142
Figure 54	Answers to the survey question number 6 . . . . .	142
Figure 55	Answers to the survey question number 7 . . . . .	143
Figure 56	Answers to the survey question number 8 . . . . .	143

# LIST OF TABLES

Table 1	Wikipedia versions in Italian dialects . . . . .	28
Table 2	Total amount of data gathered with Mercato Linguistico and Poker Parole . . . . .	80
Table 3	Geographical assessments for a phrase in Mercato Linguistico	86
Table 4	Total amount of data gathered with Termina . . . . .	106
Table 5	Classification of associated terms for the concept “map func- tion” . . . . .	108
Table 6	Differences between Agora’s data schema and the database .	121
Table 7	Classification of associated terms for the concept “map Funk- tion” (in German) . . . . .	139



## Part I

# INTRODUCTION AND RELATED WORK



# 1 | INTRODUCTION

Crowdsourcing is a neologism created by Jeff Howe in 2006 that denotes an open call for participation to an undefined group of people in order to complete work that is traditionally performed by a designated agent [249]. Such a designated agent is, e.g., an employee getting paid by a company or a volunteer contributing to a non-governmental organisation. Similarly, the manifestations of crowdsourcing range from paid tasks over tasks with an uncertain probability of a payout to unpaid, voluntary contributions. The incentives for participants are quite different, the tasks are manifold, and the quality of data is an issue, making it challenging to design the right application for a specific task. Such issues are increasingly being addressed in scientific research whereupon crowdsourcing has started to become a major direction for research, bringing together different areas of research like computer science, psychology, biology, pedagogy, art history, and linguistics [8, 10, 25, 64, 149].

In this thesis, crowdsourcing techniques are investigated within two different areas: empirical research in sciences oriented on humans –focusing on linguistic field research– and e-learning. The research methodology used in this thesis follows the design-science paradigm [79]. This paradigm seeks to create innovations as artifacts, e.g., information system conceptualisations, organisational structures, social systems, or training, for solving business-relevant problems. The paradigm is complementary to the behavioural-science paradigm which aims at developing and verifying theories explaining human or organisational behaviour [79]. Together, both paradigms contribute to research in the information systems discipline. In design science, artifacts are designed, built, and evaluated iteratively to gain knowledge and understanding of a problem domain and its solution [79].

Traditionally, in linguistic field research scientists are sent to the speakers' locations for gathering data and meta-data needed for linguistic field research. They interview speakers on-site, record and transliterate the interviews, and report on these interviews by filling in forms. This process is time-consuming, costly, and possibly biased. Crowdsourcing may be helpful to solve these problems. One aim of this thesis is to investigate how crowdsourcing techniques can be employed to speed up field research, make it less expensive, and prevent biases. Therefore, this thesis introduces the artifact of a generic operating system, called Agora, for running market-based crowdsourcing applications. In applications built with Agora, a community of users can share symbolic goods, characterise properties of interest of these symbolic goods, create assessments of such characteristics, and trade with these assessments. Symbolic goods can be text, audio, or video. An assessment contains a set of characteristics and an estimation what percentage of users of the application specify the same characteristics for the symbolic good as the user herself does. Such assessments have

a monetary value associated with them representing the correspondence to the current community opinion on the symbolic good.

Mercato Linguistico and Poker Parole, two social media applications built with Agora have been conceived as part of this thesis for gathering data for Italian linguistic field research. These two artifacts are deployed on a platform called metropolitalia [24, 260], which was set up specifically for these two applications. In both, the symbolic goods under research are written phrases in Italian language varieties. As characteristics, the geographical region in which they are used, the speakers' social attributes age, gender, and level of education, and the specification of linguistically relevant words are gathered. Assessments on geographical characteristics collect perception data about users' estimations of the spread of their specified geographical characterisation. In Mercato Linguistico mainly phrases with widely acknowledged linguistic traits are collected, whereas Poker Parole focuses on gathering phrases with relatively unknown characteristics. Therefore, the two applications complement each other in the data they gather. An evaluation of the applications shows details about their usage and exhibits their potential to gather high quality data useful for linguistic field research. Furthermore, Agora's principle of assessments can provide more precise distributions than usual characterisations. Of course, Agora is generic and it can be used in other contexts for empirical research, for example in the area of art history.

For e-learning, the second area of research addressed in this thesis, a game-like e-learning tool, called Termina, has been developed as artifact. Termina aims at supporting students as well as lecturers in university courses with many participants. In Termina, lecturers define a set of concepts for a certain topic, like a university course, and students playfully state associated terms for the given concepts. Lecturers then classify the terms into close and far, providing feedback to students. This facilitates association-based learning [180], which can support students' usual learning processes. Lecturers benefit by knowing better when misunderstandings arise and which topics are better understood than others. Together, students and lecturers contribute to the construction of association maps. This thesis proposes the name "association map" for a simplified version of a concept map, in which concepts are nodes and relationships are links between them. Association maps can support active, meaningful learning and therefore can incite students to think about topics and to gain knowledge that is interconnected with existing knowledge the students have. In an evaluation, Termina's use along a university course is studied. The motivation of the students shows that the fundamental structures of Termina are well received. The results of the evaluation affirm Termina's utility for students as well as for lecturers.

## 1.1 CONTRIBUTIONS

To sum up, the contributions of this thesis are as follows:



- A literature review over the emerging area of collective intelligence: The area of collective intelligence together with social computing, crowdsourcing, human computation, games with a purpose, and prediction markets is described, the areas are differentiated from each other, and links between the concepts are established.
- A model for a market-based operating system “Agora” for gathering data in crowdsourcing applications: Its generic concepts facilitate building applications gathering complementary and rich data and furthermore can be used in different areas like linguistics and art history.
- Two complementary social media applications for Italian linguistic field research built with Agora and results from their evaluation: “Mercato Linguistico” and “Poker Parole” have been conceived for gathering complementary data and meta-data of Italian language varieties. The quantitative and qualitative evaluation shows that Agora’s approach is convincing and that it can provide high quality data.
- A game-like e-learning tool “Termina” and results from its evaluation: Termina supports students with association-based learning and it helps lecturers to identify students’ misunderstandings. From data gathered with Termina association maps can be generated automatically. Students perceive Termina as helpful, supportive tool and lecturers profit from its benefits.
- The software architecture of the presented applications: The main principles of the Seam Framework, on which the applications are based, are described, and the modularity of the source code, the database structure, and the search functionality are explained.

In addition, the practical output of this research is as follows:

- “metropolitalia”, a platform for Italian linguistic field research: The platform is gathering data at <http://www.metropolitalia.org> for Italian linguistic field research since August 2012.
- An implementation of Termina available for courses: Termina is available for two courses at <http://termina.pms.ifi.lmu.de> for playing.
- The software for all aforementioned applications: The source code is available under the open source GNU Affero General Public License (AGPL) at <https://github.com/play4science/gwap>.

## 1.2 STRUCTURE OF THIS THESIS

This thesis is structured in five parts and ten chapters. After this introduction, [Chapter 2](#) continues [Part I](#) by summarising research related to this thesis: The subareas of collective intelligence together with its applications are analysed,

helping with the classification of the crowdsourcing applications built with Agora. Also, gaming techniques and e-learning, relevant for [Part III](#) of this thesis, are described.

[Part II](#) explains the contributions in the area of linguistic field research in detail. Therein, [Chapter 3](#) describes the goals of the platform metropolitalia and gives an introduction to linguistic field research, especially on the Italian language, which differs from other languages in its current reorganisation. [Chapter 4](#) introduces the market-based operating system Agora with its core concepts, it describes the role of play-money in Agora, the user interactions, Agora's similarity to a financial market, and means for ensuring the quality of the gathered data. This sets the basis for the next chapter. [Chapter 5](#) describes the two complementary crowdsourcing applications built with Agora, Mercato Linguistico and Poker Parole, explains user incentives, the cold start problem, the applications' classification in the field of collective intelligence, and Agora's potential in art history. The applications thus demonstrate Agora's features. [Chapter 6](#) provides an evaluation of the data gathered with Mercato Linguistico and Poker Parole, including the temporal development of the gathered data, a quantitative breakdown of numbers, an evaluation of the assessment's estimated agreement proportion, and the quality of the gathered data for linguistic field research. This evaluation yields insights into the usage of the applications.

[Part III](#) focuses on e-learning and specifically on the game-like e-learning tool Termina. [Chapter 7](#) introduces the concepts of Termina, its modes of operation, and its benefits for students and lecturers; it explains how association maps are constructed and how Termina can be employed for expert communities. After the description of Termina, [Chapter 8](#) gives an evaluation of Termina along a university course, evaluating the quantity and quality of the gathered data, showing the results of a user survey, and giving brief concluding remarks.

[Part IV](#) and its single [Chapter 9](#) give an overview over the software architecture of the presented applications, starting with an introduction to the Seam Framework and the JBoss Application Server and continuing with the descriptions of the modular concept of the project, the database structure, and the search with Solr. These information are relevant for researchers who want to extend one of the presented applications.

Finally, [Part V](#) concludes this thesis with [Chapter 10](#), summarising the main findings of this thesis and giving an outlook to future work.

## 2 | RELATED WORK

Aspects of the content of this chapter have been published in [24, 25, 107–111, 130]: [Section 2.1.2](#) extends parts of [24], [Section 2.1.3](#) is partly based on [24, 130], [Section 2.1.4](#) is derived from [24, 25, 107, 110, 111, 130], [Section 2.1.5](#) is based on [24, 107, 110, 111, 130], [Section 2.2.1](#) is derived from [24, 107, 109–111], [Section 2.3.2](#) extends parts of [110], and [Section 2.4](#) is based on [108].

This thesis mainly builds upon research results from the field of collective intelligence, which comprises among others crowdsourcing and human computation. This research field is introduced in the successive [Section 2.1](#), collective intelligence applications are presented in [Section 2.2](#). In the subsequent sections, serious games and gamification are covered as well as e-learning, a field of relevance for the second part of this thesis.

### 2.1 COLLECTIVE INTELLIGENCE

Collective intelligence is a broad area. This field is concerned with “groups of individuals doing things collectively that seem intelligent”, as defined by Malone, Laubacher and Dellarocas [137]. Smith [188] emphasises the emergence of intelligent behaviour and defines collective intelligence as the concept that “a group of human beings can carry out a task as if the group, itself, were a coherent, intelligent organism working with one mind, rather than a collection of independent agents”.

As the name collective intelligence and its definitions suggest, it is a broad area and collective intelligence in this sense has been around in groups like families or village communities for a long time. The performance of human groups for tasks like brainstorming, solving visual puzzles, making judgements, and negotiating over limited resources has, for example, been studied [221]. The result is that social sensitivity of members of the group is a significant factor for the emergence of collective intelligence, which cannot be explained by the individual intelligence of the group members [221]. Also among animals collective intelligence is considered to exist. E.g., social insect colonies like ant colonies make decisions collaboratively and with sophisticated choices [61, 136]. Collective intelligence is also used to describe the behaviour of a group of physical robots interacting with each other [140]. Lévy [131] does not consider the intelligent behaviour of ant colonies as collective intelligence, because ants do not have a vision of the whole and are individually unintelligent. He specifies collective intelligence as an intelligence that is distributed everywhere, that continuously creates its value, that is coordinated in real time, and that can mobilise competences effectively [131]. The rise of the Web brought human

collective intelligence to a new level. Collaboration can happen on a global scale, much faster, and more efficient than in traditional collective intelligence [131]. In contrast to the more philosophical view of Lévy, this thesis is concerned with Web-based collective intelligence, as in the definition of Malone, Laubacher and Dellarocas [137].

Malone, Laubacher and Dellarocas [137] classify collective intelligence systems by so-called “genes” identified from the four questions: *Who* is performing the task? *Why* are they doing it? *What* is being accomplished? *How* is it being done? Each collective intelligence system can be classified based on these four genes. On their website [246], they provide an overview over the field of collective intelligence and an annotated list of examples. Kapetanios [99] gives an historical overview of collective intelligence, lists contributing fields of research, and state issues to be answered by future research. Several of these have been tackled in the meantime, for example, how to cope with uncertainty and probability in data management (see, e.g., [153, 201, 208]) and how learning can be enhanced with computer games (gamification, see Section 2.3.2).

How and in what areas collective intelligence impacts companies and their business decisions is investigated by Bonabeau [20]. He concludes that problems in companies may be solved by a “group of diverse, independent, and reasonably informed people” [20] better than by individuals. However, the development of the right tool for a certain problem is a tricky task.

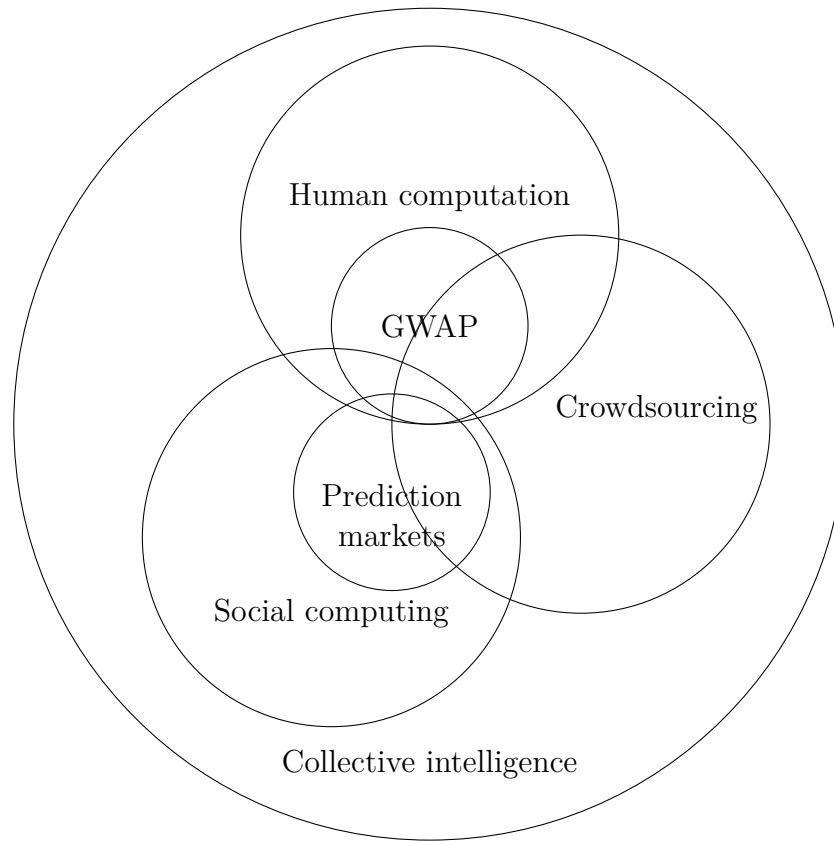
Quinn and Bederson [166] propose a well received taxonomy of the field of collective intelligence and its associated fields human computation, crowdsourcing, social computing, and data mining. These terms are defined based on previous research and a classification is suggested based on six distinguishing factors. Some factors are very similar to the ones in [137], whereas others are added, adapted, or removed.

Figure 1 illustrates the field of collective intelligence and related fields relevant for this thesis. The figure is based on research by Quinn and Bederson [166]. In addition, games with a purpose (GWAPs) and prediction markets (see Section 2.1.4 respective Section 2.1.5) are included at appropriate places, while data mining has been removed because it does not play a primary role for this thesis. The field of human computation has been moved so as to fully belong to collective intelligence, as it is explained in Section 2.1.3. The exact definitions of the covered fields are given in the subsequent sections.

In the following, the areas of social computing, crowdsourcing, human computation, games with a purpose, and prediction markets are discussed.

### 2.1.1 Social Computing

Social computing is defined by Erickson [58] as the use of digital systems that support social interaction, first and foremost by providing communication mechanisms but also by recording individual’s actions and interactions, processing them, and displaying them for further interaction. Along the lines are definitions of other researchers [54, 160, 178]. According to this definition, email, instant messaging, photo sharing, weblogs [118], wikis, social games,



**Figure 1:** Collective intelligence and related fields depicted in a diagram with (partly) overlapping regions. Adapted and enhanced from Quinn and Bederson [166]. The definitions are given in the respective sections.

and virtual worlds [144] belong to social computing. Besides the use of digital systems for the online world, e.g., mixed reality systems for interactive games [35] and pervasive systems facilitating social interaction [54] belong to social computing.

According to these definitions, the focus of social computing is on facilitating communication and interaction between humans. Recent research, however, also considers social studies and simulation techniques as belonging to social computing [209, 226]. Including these areas into social computing can help to improve social computing applications. This is done, for example, in the research area of social signal processing, which “aims at providing computers with the ability to sense and understand human social signals” [202], like attention or politeness. Therefore, in this thesis the broader definition including social studies is used.

Parameswaran and Whinston [159] give an overview over the broad field of research relevant for and impacting social computing. The cross-disciplinary theoretical research and the infrastructure underlying social computing are outlined by Wang, Zeng, Carley and Mao [209]. Computational approaches in the area of social computing are examined by King, Li and Chan [103].

Terms often mentioned in the context of social computing are social software and social media. The origin of the term social software is described

by Allen [230]. Shirky [269] defines social software as software that supports group interaction. Coates defines it as concerned “with the augmentation of human social and / or collaborative abilities through structured mediation” [236]. Definitions vary but all include software facilitating interaction, often—but not necessarily—based on the Internet. Thus, social software includes, e.g., wiki or blog software. Warr [210] gives more examples and an analysis of the prospects of social software. Social software thus denominates the tools used for social computing, whereas social computing is a paradigm.

Social media, according to Kaplan and Haenlein [100], are Internet-based applications, the users of which can continuously participate and collaboratively create content, which is then shared and exchanged. Social media are seen as opposed to traditional media in which content is created and published by individuals acting alone. Compared to social software, which focuses on the software part, social media focuses on the media itself and its user interaction and collaboration.

Social computing partly belongs to collective intelligence because social interaction is a kind of intelligence. Collective intelligence is more general than social computing in that it encompasses also applications without direct social interaction, for example, some human computation applications (see [Section 2.1.3](#)).

### 2.1.2 Crowdsourcing

The term crowdsourcing has been mentioned first by Howe [249] and derived from the term outsourcing. He defines crowdsourcing as “the act of taking a job traditionally performed by a designated agent (usually an employee) and outsourcing it to an undefined, generally large group of people in the form of an open call” [248]. In the words of Doan, Ramakrishnan and Halevy [53], crowdsourcing denotes the participation of many humans on the Web to solve a common problem that is defined by the system owner. This broader, widely adopted definition of crowdsourcing does not specify that the type of problem is traditionally performed by a human, as Howe does. Quinn and Bederson [166], like Howe, want to keep the focus on this more specific problem domain, in which the job is traditionally performed by a human, in order to distinguish crowdsourcing better from other terms like human computation. The nature of collaboration can be explicit or implicit [53]. In explicit crowdsourcing contributors know what they contribute to and usually know the goal of the system. In implicit crowdsourcing users do not need to know the goal which is specified by the system owner.

Crowdsourcing is applied in many different contexts like the collaborative web platforms Wikipedia [275] and Stack Overflow [270], paid user studies [104], selling user-generated content [21], judging the relevance of documents [6], ideas competition [127], answering complex queries [60], or disaster response [65, 227]. Also for companies, crowdsourcing is a valuable field [207]. The open source movement is cited as a successful incarnation of crowdsourcing [5,

[53], though it differs from crowdsourcing because in open source everything is published whereas in crowdsourcing the results of the collaborative process are often not published [21]. Open source can thus be considered more liberal than crowdsourcing. Additionally, crowdsourcing is applied in human computation, games with a purpose, and prediction markets, as described in Sections 2.1.3, 2.1.4, and 2.1.5. Overviews of crowdsourcing systems are given in [53, 224].

Crowdsourcing is often implemented by means of online labor markets like *Amazon Mechanical Turk (AMT)* [231]. AMT is an online labor market in which providers set up small tasks, also called *human intelligence tasks* (HITs); users of AMT can solve the tasks as workers and are paid for completing these tasks [189]. AMT is, for example, used for user studies [104]. In general, data generated through AMT is more diverse but also of lower quality than expert data [189]. To compensate for the lower quality the same HIT is often given to several workers. As the costs for paying AMT workers are lower than paying a regular worker, e.g., a student at university, it can be cheaper to crowdsource a task on AMT than to solve it employing a traditional worker. Furthermore, solving tasks using AMT may be faster or more responsive, especially for tasks that need to be processed immediately. On-demand, real-time crowdsourcing through AMT is available as well [15, 120]. Frameworks for coordinating complex and inter-dependent tasks exist that help constructing applications for AMT [105, 133, 146, 147]. The payments to workers have been investigated and strategies for optimal pricing suggested [84, 86, 139, 147]. The quality of data gathered via AMT is mostly good when techniques like employing several workers, analysing the noise level, uncertainty, and ambiguity are employed [4, 88, 93, 174, 189]. Research on rewards for users on paid crowdsourcing platforms shows that the return of rewards correlates logarithmically with participation level, i.e., a higher participation does not yield a corresponding increase in return [52]. The high amount of research on AMT led to criticism of research on applications of AMT in which no new findings are gained [3]. Instead, research should focus on improving crowdsourcing methods [3]. Furthermore, an analysis of seven crowd work platforms shows alternatives to AMT, offering solutions to limitations of AMT [198].

Crowdsourcing clearly belongs to collective intelligence, because in crowdsourcing, several humans are engaged to solve a problem, which is a way of collaborating in a way that seems intelligent. However, collective intelligence is more general than crowdsourcing: in collective intelligence no open call needs to be made and the nature of the problem is unrestricted. While in crowdsourcing the problem to be solved is traditionally performed by a designated agent and then outsourced, the problem may exist implicitly in collective intelligence.

Several crowdsourcing applications incorporate social computing techniques, especially explicit ones in which users collaboratively help to solve a problem they are familiar with and use social interaction in the problem solving process. Implicit crowdsourcing applications that gather and evaluate user data without social interaction of users should be considered as not being part of social computing. Quinn and Bederson [166] furthermore highlight that the foci of the terms crowdsourcing and social computing are different: while crowdsourcing



aims at the collaboration for solving problems, social computing facilitates human interaction. It should be noted that, nevertheless, the distinction between social computing and crowdsourcing is not sharp and a clear division is not meaningful.

### 2.1.3 Human Computation

Human computation has been carried out long before computers existed. For example, the calculation of the trajectory of the Halley's Comet in 1758 was performed by three highly skilled astronomers jointly, as the calculation was too comprehensive for a single astronomer [68, 123]. The modern usage of the term human computation however –which is of interest here– has been introduced by von Ahn [204] in his PhD thesis as a paradigm for taking advantage of human processing power in order to solve problems that so far cannot be solved by computers. Definitions of other researchers vary in the scope but are compatible with each other and agree on the following two aspects of human computation: the problem might some day be solvable by computers, and a computational system or process organises human participation [166].

Law and von Ahn [123] highlight two aspects they consider necessary for human computation: consciousness and explicit control. First, humans must have a *conscious* role in determining the outcome of the computation [123]. Thus, they consider projects in which humans do not perform computation themselves but merely provide their computers' processing power –like in the projects Folding@home [242] or SETI@home [267]– as not belonging to the area of human computation. For participatory sensing projects [28] they distinguish between human computation where active decisions take place and general collective intelligence where humans act as passive sensor carriers with no possibility of an active choice. Second, the level of *explicit control* must be greater in human computation than in general collective intelligence [123]. The process of distributing tasks to humans and combining human answers therefore happens in an organised, well-defined manner, similar to an algorithm specifying what gets processed by whom and how [123].

In human computation algorithms, the problem one wants to solve is usually broken down into smaller tasks that are routed to computers or humans for being solved [123, 132]. Then, individual outputs are aggregated, which can also be done by computers as well as humans [123]. This leaves much room for developing different scenarios how humans and computers work together.

Quinn and Bederson [166] position human computation as not fully belonging to collective intelligence. They argue that a human computation approach is conceivable that includes a single human worker, though they admit that such a system does not exist yet. Law and von Ahn [123] in contrast see human computation as fully belonging to the area of collective intelligence because a participation of multiple humans is the strength of human computation, solving problems of inaccuracy, lack of expertise, untruthful answering, cheating, and lack of diversity. Also Parameswaran and Polyzotis [158] see the need for



employing several humans because of the uncertainty involved when trusting a single human. Therefore, human computation is considered a part of collective intelligence in this thesis.

Common with social computing are approaches in which humans interact with each other in a social way. In contrast to social computing, social interaction is not necessary for every human computation system and only some social computing systems involve human computation. Quinn and Bederson furthermore distinguish human computation and social computing through their foci: social computing facilitates relatively normal human behaviour through technology and human computation tries to solve computational problems with the help of humans [166].

Human computation and crowdsourcing share approaches in which tasks that are (in principle) computable are outsourced to a crowd of humans consciously engaging under explicit control. In contrast to crowdsourcing, human computation can happen with humans recruited traditionally or not in the form of an open call [123]. Quinn and Bederson distinguish human computation and crowdsourcing in that “whereas human computation replaces computers with humans, crowdsourcing replaces traditional human workers with members of the public” [166].

To summarise, the field of human computation belongs to collective intelligence, intersects with the fields of social computing and crowdsourcing, and also has a separated part (as depicted in Figure 1).

#### 2.1.4 Games With a Purpose (GWAPs)

A special form of human computation are human computation games, also called *games with a purpose* (GWAPs) [203, 205, 206]. Von Ahn and Dabbish [206] introduced the term GWAP for describing a game which is designed such that many users together solve a given problem while playing the game. The first GWAP, called the *ESP game*, solves the image labeling problem, in which digital images need to be annotated with labels, e. g., to be found by text-based search engines. In the ESP game, the same image is shown to two randomly paired users who are rewarded if they suggest the same label for that image. Labels (also called tags) which have been verified for that image and do not have to be entered by users again are displayed as so-called taboo tags. Since the only resource shared by the two users is the image, the users tend to enter descriptions that are likely to be given also by their counterparty user. Thus, images are labeled with descriptions while users are playing the game [206]. The ESP game has been criticised for displaying taboo tags to users because it is argued that users tend to enter similar tags [19]. An algorithm has been devised that successfully plays the ESP game using a synonym dictionary [211]. But this only affects the game design of the ESP game itself and not other GWAPs.

Several other GWAPs have been designed that solve different problems, amongst others games for the semantic web [185], for biosciences [43, 64], for geo-spatial data [141], for social networks [168], for common sense data [191],

for lexicons [212], and for linguistics (see Section 2.2.1). Also in art history, GWAPs on the “Artigo” platform [233] –developed in the Play4Science project– are employed to gather descriptive tags for artworks [25, 112, 192]. Suggestions for an extension of the ESP Game are given in [26]. A survey on games for knowledge acquisition is given by Thaler, Simperl, Siorpaes and Hofer [195]. Pe-Than, Goh and Lee [196] classify human computation games using a typology consisting of 12 dimensions.

The mechanisms for aggregating the individual outputs in a GWAP are categorised by the manner how an agreement is reached. The mechanisms of output-agreement [205], input-agreement [124], complementary-agreement [122], and function computation [123] are distinguished: In output-agreement, the same input is shown to two players, and an agreement is reached if both specify the same output. In input-agreement, either the same input or two different inputs are shown to two players, they can communicate, and an agreement is reached if both specify that the input is the same or if both specify that the inputs are different. In complementary-agreement, two players get the same input and one player has to select “positive” output, the other “negative” output, and an agreement is reached if they specify complementary outputs, i.e., there is no overlap. In function computation, players must perform a task before being able to compute the actual agreement outcome. It is not disjunct with the other agreement mechanisms. For example, the input-agreement mechanism described above belongs to function computation. Ho and Chen [83] distinguish simultaneous and sequential verification games and model these using game theory. The game theory in GWAPs is further investigated by Huang and Fu [89], Jain and Parkes [94] and Jain and Parkes [95].

Besides these agreement methods, further research focuses on other scoring techniques for getting truthful responses from players if the validity of the output is unknown [208]. In the peer-prediction method, players are rewarded based on the difference of their output to a predicted output, which is calculated by the system based on previously stated responses [145]. This eliminates the scoring directly on a player’s output and thus can yield honest responses. However, the mechanism designer needs to know the players’ information structure and the prior probabilities for all actions, which is difficult to achieve for real applications [208]. Adjustments of this method lower the amount of information needed, but cannot eliminate it [216, 217]. The Bayesian Truth Serum (BTS) [164] does not have these prerequisites. In the BTS method, players state their output and a prediction of the distribution of responses. They are rewarded for stating “surprisingly common” outputs, i.e., responses that are more common than collectively predicted. This takes Bayesian inferences into account and can induce truthful responses. However, BTS depends on a large number of players [215]. A robust variant of BTS also works for three players, but only for binary output signals, e.g., yes / no answers [215]. So a solution for all cases is still not available and is difficult to achieve [208].

Regarding their relationships to the other fields of collective intelligence, GWAPs are a subfield of human computation and collective intelligence be-

cause GWAPs focus on human computation approaches which are played in an enjoyable, playful way by many humans.

Similar to the relations between human computation, social computing and crowdsourcing, GWAPs intersect the latter two fields. Crowdsourcing and GWAPs are obviously intersecting where the game is built as an open call for participation. However, GWAPs may exist that can be played by a well-defined amount of people and that are not crowdsourcing approaches with an open call. Furthermore, GWAPs may have social interaction but do not have to.

#### 2.1.5 Prediction Markets

Prediction markets are employed for estimating what the results of unknown future events are. Wolfers and Zitzewitz [219] define prediction markets as markets in which users trade in contracts whose payoff depends on unknown future events. Such events happen some time in the future and can, for example, be presidential elections [14] or volume of sales [33]. Important is that the future events can be verified at some point in time.

The idea of prediction markets is based on the efficient market hypothesis, which claims that in a financial market the price of an asset reflects all available relevant information [97]. Prediction markets are supposed to be efficient markets, similar to financial markets. Thus, the price of a contract correlates with the probability of the future event, which has been confirmed by various research [219, 220]. For example, prediction markets are successful in elections and also outperform polls impressively [13, 14]. Internal prediction markets have been employed in several companies, also by Google on questions like “How many users will Gmail have?” [45]. Investigations on the data reveal correlations between co-workers and optimistic biases for new employees and when Google’s stock price increased [45].

Prediction markets are also called information markets [1, 34, 73, 200] or idea futures [76, 161]. Chen, Chu, Mullen and Pennock [34] compare information markets with opinion pools with linear and logarithmic agreement functions, which are seen as equally successful with different advantages and disadvantages.

Qiu, Rui and Whinston [165] study the forecast efficiency in prediction markets when information exchange happens in social network-embedded prediction markets. Results are that social networks can improve as well as worsen the forecast efficiency, depending on the cost of information acquisition, number of friends, and network density. Furthermore, the impact of prediction markets and how the outcomes may be used for decision support have been studied in [12, 190]. Servan-Schreiber, Wolfers, Pennock and Galebach [179] show that prediction markets run with play-money perform as well as those with real money. Legal issues –prediction markets are seen as Internet gambling, which is prohibited in several U.S. states– are discussed and proposals for developing a safe legal basis for prediction markets are given by Arrow, Forsythe, Gorham, Hahn, Hanson, Ledyard, Levmore, Litan, Milgrom, Nelson, Neumann, Ottaviani, Schelling, Shiller, Smith, Snowberg, Sunstein, Tetlock, Tetlock, Varian,

Wolfers and Zitzewitz [7]. Wolfers and Zitzewitz [218] discuss five open questions in prediction markets: how to attract uninformed traders, how to tradeoff interest and contractability, how to limit manipulation, whether markets are well calibrated on small probabilities, and how to separate correlation from causation.

Decision markets are similar to prediction markets with the difference that no external event is predicted but a decision assessed [156]. For example, companies can let employees assess business decisions, e. g., whether a software prototype should be promoted to a project with a higher budget. The outcome of such decisions can not always be measured in decision markets, which makes a verification of results more difficult than in prediction markets in which at some point in time the outcome of a future event is known. Leutenmayr, Bry, Schiebler and Brodbeck [129] investigate perturbations and equilibria of such decision markets.

Concerning the related collective intelligence fields, prediction markets clearly belong to social computing because trading in a market is an action involving social interaction.

Furthermore, some applications of prediction markets are employed as an open call, thus being a form of crowdsourcing, and some are provided to a defined amount of people. Therefore, both fields intersect each other.

Usually it is not possible to compute the probability of an unknown future event using an algorithm. But there may be situations in which an algorithm can predict the future event. This is, for example, conceivable for predicting sales on the basis of some other company and market data. Therefore, certain prediction markets can be a form of human computation. Also GWAPs that employ prediction markets similarly are conceivable.

## 2.2 COLLECTIVE INTELLIGENCE APPLICATIONS

Now that the relevant subfields of collective intelligence have been identified, applications in these fields focusing on specific topics will be investigated. The topics under survey are linguistic applications, market-based applications, and applications for learning.

### 2.2.1 Linguistic Applications

Crowdsourcing has already been applied successfully in linguistic applications, mainly in theoretical linguistics. Several linguistic projects employ crowdsourcing platforms like Amazon Mechanical Turk (AMT) for data gathering. Munro, Bethard, Kuperman, Lai, Melnick, Potts, Schnoebelen and Tily [149] present several linguistic projects exploiting human computation, specifically, on AMT. An important conclusion of this article is that the linguistic quality achieved using human computation is comparable to that of controlled laboratory studies. The majority of linguistic research relies on mechanised labour, like that

AMT provides, for gathering data [173]. For example, Arabic dialects have been gathered via AMT to improve machine translation [225]. Bernstein, Little, Miller, Hartmann, Ackerman, Karger, Crowell and Panovich [16] present a crowd-powered word processing tool for proofreading, shortening, and editing texts via AMT workers. Minder and Bernstein [146] propose a general-purpose framework for human computation tasks and evaluate it by creating a translation tool from German to English.

Further articles report on using GWAPs for gathering data: Lafourcade [119] created a GWAP called *JeuxDeMots* [254] for collecting a lexical network of related terms. In a later version of *JeuxDeMots* ontological relations are additionally collected for generating typed relations. Poesio, Chamberlain, Kruschwitz, Robaldo and Ducceschi [163] employ a GWAP called *Phrase Detectives* [262] for anaphoric co-reference data in English and Italian. Players perform two tasks: providing judgements and validating judgements of other players. They are rewarded personally (with scores and levels), socially (competition with other players), and financially (small prices) for their participation. Chamberlain, Fort, Kruschwitz, Lafourcade and Poesio [30] reflect on the success of these two games and testify them promising results.

Further GWAPs exist for gathering co-reference data [82], paraphrases [38], transcriptions [157], and emotions, intentions, and attitudes [162].

In addition to active participation of users, passive, observation-based approaches to analysing social media for linguistics are investigated. For example, geotagged Twitter messages are gathered, automatically categorised into topics, and the geographical distribution of all terms measured, resulting in a geographical mapping of certain dialect terms [57]. Such passive approaches can be categorised as collective intelligence applications and –if social interaction is involved– as social computing applications, but do not belong to any other subordinate area mentioned in [Section 2.1](#).

### 2.2.2 Market-Based Applications

Applications based on a market mainly reside in the area of prediction markets. One of the earliest attempts to predict unknown future events with a prediction market is the Iowa Political Stock Market [59]. The predictions about results of presidential elections were more accurate than opinion polls in nearly all cases. Nagar and Malone [150] improve predictions with hybrid markets, i. e., prediction markets including both humans and computer algorithms, for the case of the results of football games.

Further applications of prediction markets on the Internet not only for means of research include Intrade [251], a pure prediction market which has been shut down because of legislation difficulties in 2013, and Lumenogic [256], a company offering –amongst others– forecasts through prediction markets. Similar to Lumenogic, CrowdWorx [238] offers so-called social decision support systems in which employees are asked for their opinion on a certain topic and the company receives the condensed results. Here however, the concrete algorithm is not published.

Market-based applications that are no decision markets have been proposed: Hsieh and Counts [85] designed and evaluated a market-based real-time *question and answer* system. Users promise a financial reward –in terms of virtual money– to the user with the best answer to the question they pose. Compared to the same system without a market in place, the market-based system is seen as more serious by users with less ignorant questions posed but also less social community aspect. A similar market for crowdsourcing tasks is examined by Yang, Adamic and Ackerman [222]. Research on two-sided markets, in which two different types of users interact through one or more platforms or mediators, shows that similar platforms have different characteristics in terms of cross-side and same-side network effects, e.g., in the question and answer community Stack Overflow [270] questioners grow quadratic whereas answerers only grow linearly [117].

Marketplaces for distributing crowdsourcing itself have been investigated by several researchers [87, 176, 183]. Different proposals have been made for how to bring the seller and buyer together in the most efficient way. Dynamic pricing strategies for paid crowdsourcing marketplaces have been investigated by Minder, Seuken, Bernstein and Zollinger [147].

Hsieh, Kraut, Hudson and Weber [87] propose to use market mechanisms in synchronous communication. Users asking questions pay responding users a fixed or variable price. This leads to higher productivity, but only for certain setups as incentives in systems without a market may differ from those with a market.

### 2.2.3 Applications for Learning

Collective intelligence has also been successfully applied to learning, where collaboration helps fostering knowledge acquisition. Duolingo [240] is an interesting platform, founded by von Ahn, on which people can learn foreign languages in a collaborative way. Users can choose between lessons predefined in the system (similar to online language courses) and lessons based on material from the Web (e.g., Wikipedia pages) or from customers of Duolingo who outsource translation tasks. For such texts, users translate single sentences or phrases, rate translations by other users, and improve or correct translations by other users. Duolingo helps users with the translation by displaying possible translations for single words. With the help of many users the best translation emerges. Thus, users learning a language in Duolingo at the same time participate in translating the Web. Furthermore, both computers and humans do what they can do best: Computers provide possible translations for single words as a dictionary and provide a well-designed platform for collaboration. Humans choose the fitting words and put them into the correct order for yielding a correct translation.

Besides platforms with learning as the main goal, learning as a by-product can take place as part of a participation in a crowdsourcing project. The emergence of citizen science –inspiring people of all ages to take part in scientific research– brought learning of scientific knowledge as part of certain tasks of



a project [22, 197]. For example in a bird-watching project, for which the sightings of birds are recorded together with their geographical location, participants gained knowledge about bird biology. Their attitudes toward science or the environment did not change however [22], which is also a goal of this citizen science project. Zooniverse [276] is a platform established for assembling citizen science projects on a common website. One of the platform's projects, Galaxy Zoo [243], recruits volunteers to classify galaxies and teaches how to distinguish different types of galaxies. In a study, the main motivations for participation on Galaxy Zoo were interest in astronomy, amazement by the vast size of the universe, and excitement to contribute to original scientific research [167].

Further applications focusing on learning via electronic technologies are given in Section 2.4.

## 2.3 GAMING TECHNIQUES IN STANDALONE AND COLLABORATIVE APPLICATIONS

Humans like playing games because games provide them with positive emotions, positive activity, positive experiences, and positive strengths [142]. McGonigal [142] claims that certain types of games could change the world for the better. By including elements of games into reality more closely, important global challenges could collaboratively be met.

Such positive implications of games manifest in the current trends of GWAPs (as described in Section 2.1.4), *serious games*, and *gamification*. Some such games and non-game applications exist for single users and are thus standalone, while in others users collaborate and interact with each other.

In the following, serious games and gamification are introduced. These two techniques do not necessarily belong to the area of collective intelligence, however they are also employed in certain collective intelligence applications.

### 2.3.1 Serious Games

Serious games in general are games that are not primarily designed for entertainment [2]. Most serious games proposed so far are for learning. Susi, Johannesson and Backlund [194] categorise serious games in five application areas: military games (e.g., military simulations or war games), government games (e.g., simulations for crisis management), educational games (e.g., for classroom learning), corporate games (e.g., for sales or communication training), and healthcare games (e.g., for physical fitness or cognitive functioning). Learning through such serious games is also called game-based learning [194]. Other classifications are given by Blackman [18] and Chen and Michael [32].

In addition to serious games for learning, there are also serious games for educating the user and changing her behaviour [42]. For example, Lavender [121] wants people to become more sympathetic to homeless by playing a game

in which the player takes the role of a homeless woman trying to survive on the street for 24 hours.

Serious games, which aim at learning and behaviour change, have a different focus than GWAPs, which are specifically designed to gather data while being entertaining for players. The common ground is the gaming aspect that both contain.

### 2.3.2 Gamification

A term that complements the gaming aspect from another side is gamification, a term broadly adopted in the business context for applying game design techniques to non-game experiences [244]. From a research perspective, gamification is defined by Deterding, Dixon, Khaled and Nacke [51] as the use of game-like design, elements, and characteristics in non-game contexts. For example, badges, leaderboards, or time constraints can be introduced to motivate the. A bit differently, Huotari and Hamari [92] define gamification as “a process of enhancing a service with affordances for gameful experiences in order to support user’s overall value creation”. In contrast to the first definition, the second definition disregards game elements –because no clear definition can be given what a game element is and what not– and considers gamification for both game and non-game contexts –because no clear separation between games and non-games can exist.

There are also criticisms to the trend of gamification because it seems that only the experience is changed instead of designing the game as a game [69]. This results in casual playing without much player motivation [39]. For example, Liu, Alexandrova and Nakajima [134] observed in a study that gamification incentives did not improve the performance of users, especially it did not make users participate who were not attracted by the system without gamification.

Gamification techniques are often applied in GWAPs and serious games. However, GWAPs and serious games are designed as games themselves and not as a mere bundle of game-like elements. Gamification has to be seen as a current trend that needs more research for its benefits, drawbacks, and limitations.

## 2.4 E-LEARNING

E-learning concepts are relevant for [Part III](#) of this thesis. Therefore, an introduction into e-learning, learning analytics, and concept maps is given in the following.

### 2.4.1 General Concepts of E-Learning

E-learning in general denotes the use of computer network technology in education [213].



More recently, Garrison [63] defines e-learning as “electronically mediated asynchronous or synchronous communication for the purpose of constructing and confirming knowledge. In contrast to the first, broader definition, the second focuses on the collaborative aspect. This moves e-learning away from more traditional distant learning approaches, in which pre-compiled lecture material is provided for learning on ones own. More recent advances in e-learning also speak from ”e-learning 2.0“, incorporating social features from the ”Web 2.0“ [72, 106, 171].

E-learning itself appears in many forms. There are early applications providing slides distributed for training, newer interactive computer programs, or recently massive open online courses (MOOCs) [46]. Some e-learning tools include game-like elements, i.e., gamification, for encouraging the user’s participation and improve the learning process [169].

Criticisms of e-learning argue that research and application design is mainly done from a technological standpoint and not from a pedagogical one [66, 75]. More insights into e-learning from pedagogy would be helpful for advancing e-learning.

The areas of e-learning and collective intelligence are related and have sub-areas in common. Especially more recent advances in e-learning tap into social collaboration and therefore include aspects of collective intelligence.

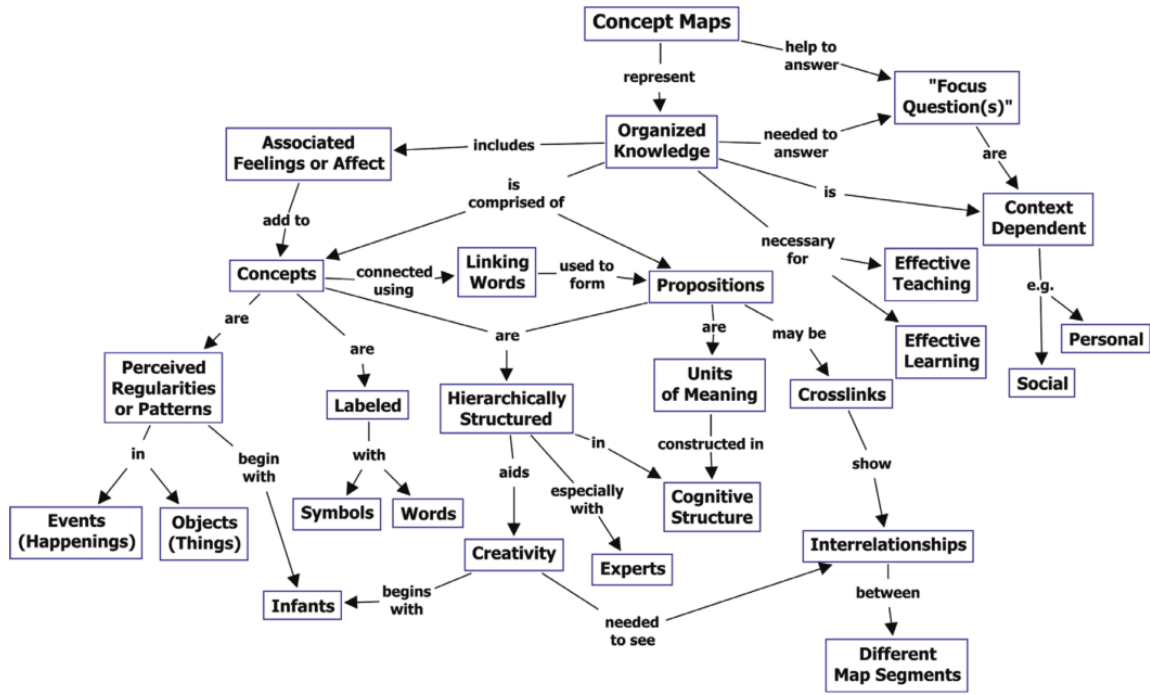
#### 2.4.2 Learning Analytics

Learning analytics is a relatively new field with the goal of understanding and improving e-learning. It developed as part of e-learning and data mining trends and the possibility to analyse large and complex datasets efficiently in short time, possibly even in real-time. Learning analytics is defined by Siemens and Long [182] as the collection and analysis of data about learners for improving learning. Such analysis gives useful insights into how learners acquire knowledge and may be used to improve learning. Duval and Verbert [56] see two major approaches of learning analytics: analysis and data mining of data generated by learners to point out patterns (see also [172]) and visualization of the data for steering the learning process (see also [55]).

#### 2.4.3 Concept Maps

Concept maps have first been developed by Novak and Dismas [152] for fostering children’s and students’ understanding of science concepts and evaluating their progress in a twelve year long study. Concept maps are graphical tools for organising and representing knowledge [151] and consist of concepts –visualised as text in boxes or circles– and relationships between them –indicated by lines– which are usually specified with one or more words. An example of a concept map is shown in Figure 2.

Concept maps are structured hierarchically with the most general concept on the top. The inclusion of cross-links is an important characteristic of concept



**Figure 2:** Concept map about concept maps. The figure is included from [29] with the authors' permission.

maps. Also, examples of concepts may be included in the map for clarification [151]. Concept maps have proven to be useful not only for evaluation but also for, e.g., brainstorming and learning [151]. An approach of employing computer-based concept mapping in schools as learning environment shows encouraging results [29]. A further study on the construction of concept maps shows that collaboratively created concept maps have more cross-links and thus are richer in information than individually created concept maps [44].

After having summarised the state of research in the field of collective intelligence, its interesting applications, gaming techniques and e-learning, the next parts describe the objects of this thesis.

## Part II

# LINGUISTIC FIELD RESEARCH



## METROPOLITALIA: A CROWDSOURCING PLATFORM FOR ITALIAN LINGUISTIC FIELD RESEARCH

Aspects of the content of this chapter have been published in [24, 107, 109–111]: [Section 3.2.1](#) is derived from [24, 107, 109–111], [Section 3.2.2](#) extends parts of [24, 107, 110], and [Section 3.3](#) is based on [24, 109–111].

This thesis proposes to conduct field research on the Web and describes metropolitalia [260], a Web platform designed for this purpose. Specifically, the platform metropolitalia has been conceived as a crowdsourcing platform for conducting linguistic field research for the Italian language. For the design-science methodology, metropolitalia serves as an artifact for evaluating the model of Agora.

The development of metropolitalia has been part of the Play4Science project [264], in which crowdsourcing techniques have been conceived and developed for exploiting existing documentary sources and generating new empirical data in the humanities. Within Play4Science, two platforms have been developed, ARTigo [233] for art history and metropolitalia for linguistic field research. ARTigo is a crowdsourcing platform for image tagging, especially for gathering tags for artworks. A data analysis of the gathered data is employed for providing a search for artworks by tags. ARTigo has been in use for several years and has proven to gather much and useful data and furthermore to yield useful insights into the perception of art history [41, 112, 192].

Metropolitalia has been conceived as a platform for linguistic field research, providing insights into the regional and social differences in the use of varieties of the Italian language. Both platforms, metropolitalia and ARTigo, share the same software basis, therefore other platforms benefit from improvements on one platform. The e-learning tool Termina (see [Part III](#)) is also built on this software basis. Details on the software architecture itself are described in [Chapter 9](#).

The name “metropolitalia” is a combination of two Italian words, “metropoli” (in English: metropolises / big cities) and “italia” (in English: Italy), standing for the current change of the Italian language in big cities. A more detailed description of these changes is given in [Section 3.2.2](#).

In the following, metropolitalia’s goals are explained, background information about linguistic field research on the Italian language is given, and the focus of metropolitalia on Italian varieties is illustrated.

### 3.1 METROPOLITALIA'S GOALS

Metropolitalia has several goals on different levels of abstraction. For the design-science methodology applied in this thesis, metropolitalia's main goal consists in providing an artifact for evaluating the market-based model of Agora. To this aim, metropolitalia has been conceived as an open platform with which *information about Italian language varieties* are gathered and provided for inspection by researchers. Data gathered include examples of phrases that show particular differences in language use as well as meta-data associated to these phrases. Meta-data can be the meaning of the phrase in standard Italian, geographical data, i. e., the region(s) to which a phrase is somehow related or where the phrase is spoken, or social data, i. e., characteristics like age, gender, or level of education of speakers who use the phrase.

On a more concrete level of abstraction, three further goals are concerned with the success of the platform itself. First, metropolitalia tries to attract native speakers of Italian (or a variety of the Italian language) to *share their own knowledge* about the collected phrases. Attracting many people is necessary in order that the platform reaches a critical mass of users and therefore sufficient data. Data input is voluntary without money being paid.

Second, metropolitalia aims at becoming a hub on which people can *look up data* about Italian language varieties. For example, words only used in certain regions may be looked up by people from other regions and their meta-data examined. This benefit of being an openly accessible database should contribute to the second goal, that is, attracting native speakers of Italian to share their own knowledge.

Finally, metropolitalia should exist and be maintained for a *long time* (many years) so that changes in the use of language can be observed and that the platform is established as a reference for look-up purposes.

### 3.2 LINGUISTIC FIELD RESEARCH ON THE ITALIAN LANGUAGE

For understanding the topic for which metropolitalia has been conceived, linguistic field research on the Italian language is briefly introduced in this section, together with the traditional way of conducting linguistic field research in general, the history of the Italian language with the language's specific properties, and reasons for humans' interest in language.

#### 3.2.1 Traditional Linguistic Field Research

Linguistic field research is concerned with gathering and analysing speech data in written and spoken form from speakers of some language(s) under observation. The gathered data comprise the speech data itself as well as characteristics of the speakers such as their geographical origin and social characteristics

like age, gender, or level of education, the situation in which the speech takes place –like formal or informal–, the time at which it takes place, and also whether the language is used in written form or spoken [115].

Traditionally, such multi-dimensional data are collected by scientists, typically doctoral students or low paid researchers, at the speakers' locations, usually in certain geographical regions, where they interview speakers, record and transliterate the interviews, and report on these interviews by filling forms. This process is time-consuming because each researcher can only interview a limited number of speakers. It is costly because the researchers or students involved have to be paid. And furthermore it can be biased. Indeed, researchers' conscious or unconscious preconceptions might affect how people are selected and how their answers are written down by the linguists conducting the field research [49, 125]. Because traditional linguistic field research is expensive and time consuming, only small scale studies focusing on rather small geographical regions or on a specific set of linguistic features are usually conducted [17].

### 3.2.2 History of the Italian Language

The summary of the history of the Italian language presented here is based on research by Clivio, Danesi and Maida-Nicol [40], Lepschy and Lepschy [128] and Mioni and Arnuzzo-Lanszweert [148].

The roots of the Italian language stem from ancient Latin and thus Italian has a Roman origin. For a long time, there has not been *one* Italian language, but several competing vernaculars. Vernaculars are varieties of a language spoken as native language in certain geographical regions, they are often, but not necessarily, dialects [128]. Dialects are associated to a standard language while vernaculars are not necessarily associated to a standard language. Beginning in the 12th century, attempts have been made to establish one variety as a standard language. In literature, Dante wrote his “Divina Commedia” in the 14th century in a Florentine dialect and set the basis for a national literary language. Other dialects competed in becoming the standard and were adopted by different authors, but only in written language. In spoken language, no national standard language existed until the Italian unification in the 19th century. The language landscape consisted of several rather disparate dialects.

During the restructuring and standardisation process of the Italian unification, a common language emerged due to efforts of important authors, amongst them Ascoli and Manzoni. The new standard language, evolving from the Florentine dialect, has been adopted by other authors. After the unification, the illiteracy dropped from 75% in 1861 to 50% in circa 1900 and to 40% in 1911 [128]. The standard language was taught, spread, understood, and spoken more and more in everyday language use. The influence of the unification on Italian language is investigated in detail by De Mauro [50].

In addition to the standard Italian that existed since then, dialects continued to be used. Their use was discouraged in the beginning and middle of the 20th century. Later, instead of being perceived as languages for less educated people, the Italian dialects gained acceptance in today's spoken and written language

Name	URL	Articles
Tarantino	<a href="http://roa-tara.wikipedia.org/">http://roa-tara.wikipedia.org/</a>	9,181
Emilian-Romagnol	<a href="http://eml.wikipedia.org/">http://eml.wikipedia.org/</a>	2,323
Friulian	<a href="http://fur.wikipedia.org/">http://fur.wikipedia.org/</a>	3,084
Ligurian	<a href="http://lij.wikipedia.org/">http://lij.wikipedia.org/</a>	3,169
Lombard	<a href="http://lmo.wikipedia.org/">http://lmo.wikipedia.org/</a>	28,938
Neapolitan	<a href="http://nap.wikipedia.org/">http://nap.wikipedia.org/</a>	13,905
Piedmontese	<a href="http://pms.wikipedia.org/">http://pms.wikipedia.org/</a>	62,709
Sardinian	<a href="http://sc.wikipedia.org/">http://sc.wikipedia.org/</a>	3,079
Sicilian	<a href="http://scn.wikipedia.org/">http://scn.wikipedia.org/</a>	23,115
Venetian	<a href="http://vec.wikipedia.org/">http://vec.wikipedia.org/</a>	10,187

**Table 1:** Wikipedia versions in Italian dialects. List from [235], number of articles from [255] as of 6 September 2013. For comparison, the English Wikipedia had 4,321,688 articles on this date.

across all social groups [114]. A witness of the strength of the Italian dialects is their presence on Wikipedia: There are small but lively versions of Wikipedia in about a dozen of Italian varieties (see Table 1). Currently, the varieties spoken in large Italian cities evolve further. Especially, new varieties emerge, disconnecting metropolises from one another [115].

In current Italian, a major reorganisation and restructuring of the language is happening. The new standard Italian is enriched by elements of spoken language while characteristics of these elements change during this process. New, regionally different standard languages emerge [143].

Compared to other languages like French, English, or German, the Italian language experienced the standardisation more recently and varieties are therefore much more apparent in Italian.

### 3.2.3 Humans' Interest in Language

In all cultures there is a considerable interest in language issues and in reflecting on one's own language variations. This interest arises from the following reasons.

Human language itself is very important to people because it is the main means of communication. Nearly every kind of interaction between humans depends on language, both in written and spoken form. Human language exists for at least 40,000 years [98] and has evolved manifold since then.

Languages are changing through the way their speakers use words and phrases. New words are adopted, other words dismissed, and even the grammar changes. Different generations use different languages, or at least different varieties. Languages thus are "alive" and are subject to a never-ending change.

Languages furthermore distinguish social groups from others, each group having its own common language variety [154]. An own language variety strengthens the identity and the group membership. The differentiation is present for



example for ethnic groups like nations, folks, or tribes [154], and for genders [71].

The different variations of a language (or completely different languages) raise questions about the language. For example, in dialects different words are used for describing the same object, words are uttered in other ways, and specific grammatical constructions are used. Such variations are especially salient in Italian language in which distinct varieties are used for different geographical regions [115, 143].

Besides linguistic research, popular literature examines language differences and brings the topic to the people. Novels and poetry play with language and incite interest for it.

Moreover, puns are often used in literary and in spoken forms that employ certain features of language for an intended effect or for fun. Playing with language is a way of being interested in language. Also irony is treated similarly.

Many people learn foreign languages. On the one hand, the aim is to be able to communicate with people speaking a foreign language. On the other hand, interest in the foreign language and their peculiarities plays a role when someone decides to learn a foreign language.

### 3.3 FOCUS ON ITALIAN LANGUAGE VARIETIES

The problems that traditional linguistic field research encounters are especially salient with the Italian varieties:

Language varieties in general are variations of a language that are spoken by people of a certain geographical region, of a certain social group, or in a certain context. Varieties are apparent amongst others in dialects, which are local varieties, i. e., varieties spoken in a confined geographical region [77].

The manifold varieties differ from each other in vocabulary, grammar, and/or pronunciation. Some distinctive features in language use are well known in whole, or major parts of, Italy, for example the use of “bon dì” (meaning “good day”) for greeting in some valleys in South Tyrol or the use of “delizioso” (meaning “cute”) mainly by women [115]. Other distinctive features in language use are, in contrast, known only in limited parts of Italy or certain social groups.

People speaking very different Italian varieties cannot understand each other and some varieties should be considered to be separate languages because of their disparity from standard Italian [128, 148].

Another peculiarity is that people often do not know that a word is actually not standard language but instead an element of a language variety. They only get to know that the word is not standard language when speakers from another region or social group do not understand the word or use the word in a context that does not make sense to them [115].

The language used in Italy currently changes, as outlined in [Section 3.2.2](#). Therefore, it is worth studying the language varieties for a longer period of time. Few research on the current progression exists and field research data is

scarce. The enduring changes demand for data acquisition techniques adapted to long-term data gathering.

After this brief introduction of the crowdsourcing platform metropolitalia and an overview of the area of linguistic field research on the Italian language, a market-based operating system, designed for crowdsourcing applications in this area of interest, is introduced in the next chapter.

# 4

## AGORA: A MARKET-BASED OPERATING SYSTEM FOR CROWDSOURCING APPLICATIONS

Aspects of the content of this chapter have been published in [24, 110, 111]: the introduction of the chapter is partly based on [24, 111], [Section 4.1](#) extends parts of [24, 111], [Section 4.2](#) is based on [24], and [Section 4.4](#) is derived from [24, 110, 111].

The basis of the applications on the metropolitalia platform is a generic market-based operating system called Agora (Greek for “market”). It is specifically designed for gathering the rich data and meta-data needed for empirical research, and especially for linguistic field research. Following the design-science methodology, Agora is an artifact specifying a model together with methods how to apply it for applications. Based on the generic artifact Agora, market-based crowdsourcing applications can be built (see [Chapter 5](#) for two examples). To which areas of collective intelligence they belong is discussed in [Section 5.5](#).

In Agora a community of users can share symbolic goods as well as assessments of characteristics of these symbolic goods. Agora makes it possible for a user to:

- add her own symbolic goods to the market,
- propose characterisations and assessments by specifying characteristics for her own symbolic goods as well as for symbolic goods proposed by others,
- review her own assessments,
- trade assessments with other users, and
- search and browse through available symbolic goods and their characteristics.

These features have been chosen in order to collect new symbolic goods, gather characteristics of symbolic goods, include market mechanisms for fine-grained estimations of characteristics and for a game-like character, and publish the gathered data for exploration.

In applications built with Agora the symbolic goods for which meta-data are gathered can be text, audio, or video. In the case of metropolitalia, the symbolic goods are phrases, i. e., sentences or parts of sentences. For any such symbolic good, users of the application can specify characteristics of interest

for being gathered as meta-data. In metropolitalia, the geographical region and the speaker's social attributes gender, age, and level of education are characteristics. The specification of characteristics for a symbolic good is called a characterisation. The aggregation of all characterisations for a symbolic good represents the opinion of the community of users' on the symbolic good's characteristics. In addition to specifying characteristics, a user can estimate how many users of the application specify the same characteristics for the symbolic good as she does. This extension of a characterisation is called an assessment. Through assessments, the perceptions of users about characteristics are gathered for getting further insights into the users' opinions on the prevalence of the characteristics. Assessments have a monetary value associated with them that represents the correspondence to the current community opinion on the symbolic good. This monetary value changes over time while further characterisations are generated. In addition to the monetary value of assessments, users may receive dividends as rewards for their actions, which contribute to their amount of play-money. Assessments can be traded between users for a price negotiated by them. The price gives hints about the users' future valuation of the assessment. Finally, the symbolic goods can be made publicly available for being searched and browsed. Also aggregated characteristics can be displayed. This contributes to Agora's goal of being an operating system for a complete market-based crowdsourcing platform.

Agora makes it possible to build applications on the same basis, gathering complementary data. Both well-known characteristics can be gathered using assessments as well as hardly known characteristics. This is achieved for example in metropolitalia using two distinct applications gathering on the one hand the collective opinion in beauty contest-like scenarios (see [Section 5.1](#)) and on the other hand individual minority opinions in speculative scenarios (see [Section 5.2](#)).

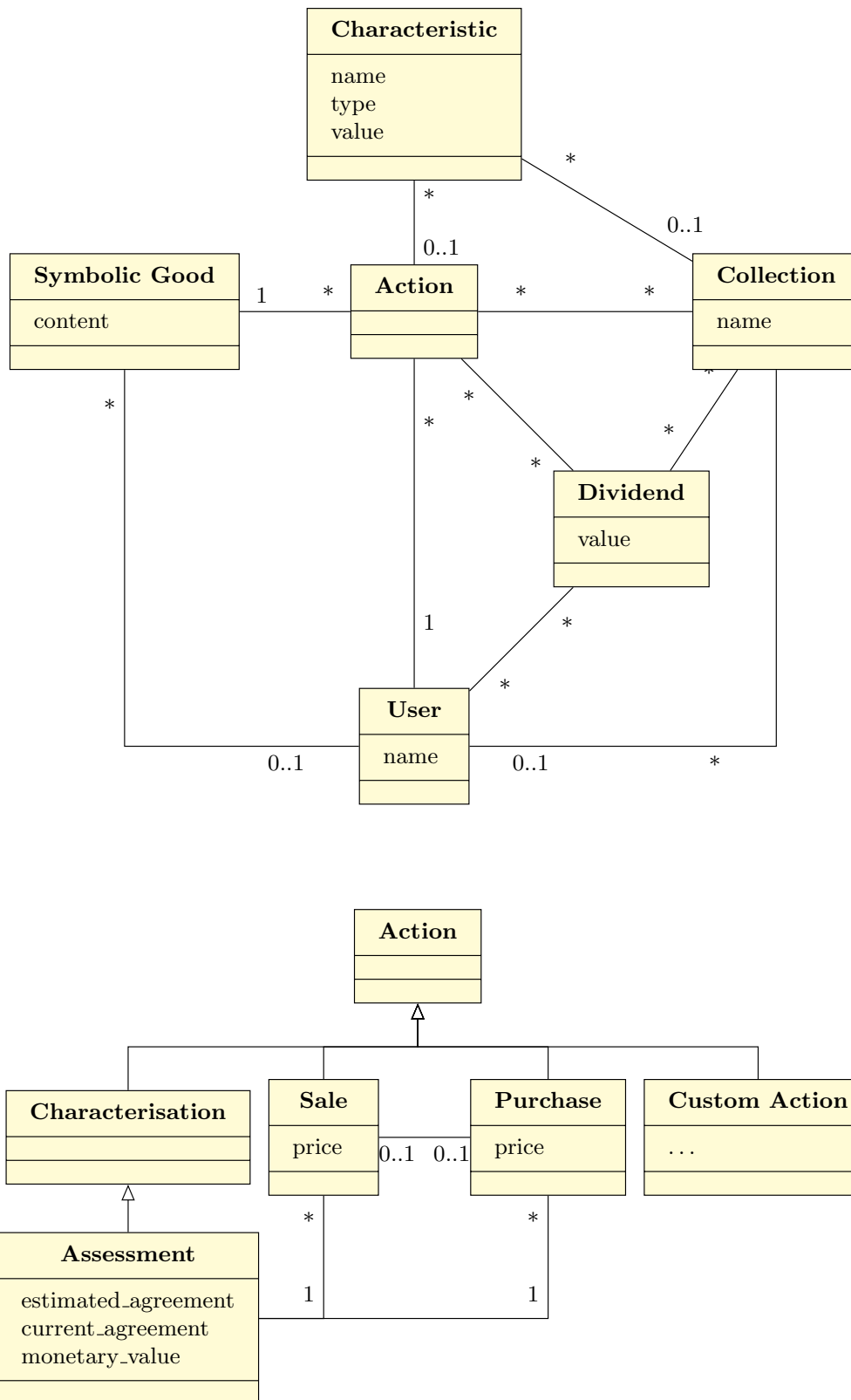
The core concepts of Agora are introduced in the following section. Next, the role of play-money in Agora is explained. Furthermore, the user interactions possible within Agora are explained, the similarities and differences of its market model to a financial market are highlighted, and measures for quality control are described.

## 4.1 CORE CONCEPTS

Agora's data schema provides the basis for understanding how data is gathered in Agora's applications. The data schema is depicted in [Figure 3](#).

Users in Agora can create symbolic goods, take actions, receive dividends, and create and own collections. Symbolic goods are the objects that are investigated in the Agora application (see [Section 4.1.1](#)). Characteristics can be specified for actions, like characterisations and assessments, and for collections. Characteristics are composed of a name, a type, and a value (see [Section 4.1.2](#)).

Actions can be performed by users and include selling and purchasing assessments, characterising symbolic goods, creating assessments, and further



**Figure 3:** Agora's data schema as a UML diagram. The classes with their main attributes without operations are displayed together with relations between them. The multiplicities of the relations are shown as numbers or as "\*", meaning any number.

custom actions defined by applications built with Agora (see bottom diagram in [Figure 3](#)). Characterisations and assessments are the core elements for gathering data and meta-data about symbolic goods in Agora. A characterisation is owned by one user, based on one symbolic good, and specifies one or more characteristics. An assessment additionally has an estimated agreement proportion. This means that an assessment represents a user's estimation of which characteristics fit to a certain symbolic good and in addition her estimation of what proportion of users agree to these characteristics. The computation of the current agreement leads to the monetary value of an assessment (see [Section 4.2.1](#)). Users can offer own assessments for sale, represented in the schema as a sale, and bid on other users' assessments, represented as a purchase. Upon agreement on a price for an assessment, the trade is completed and the respective sale and purchase are linked to each other (see [Section 4.2.3](#)). Custom actions, like revealing additional information for a symbolic good or rating characterisations or assessments, can be executed by users, also on collections. Actions are further explained in [Section 4.1.3](#).

Actions may lead to the (positive or negative) payoff of dividends to users. Dividends can be distributed to the initiator of the action or also to other users, e.g., users who characterised the same symbolic good on which an action is performed. Dividends contain the value of the payoff, i.e., play-money (see [Section 4.2.2](#)). Which dividends are distributed (if any) and what their value is has to be defined for Agora's applications.

Collections can be created by users as a way to bundle assessments with a common theme. Collections contain one or more assessments and can have characteristics.

#### 4.1.1 Symbolic Goods under Consideration

A symbolic good can be any immaterial good that can be created on the users' computers, transferred over the Internet, and displayed or replayed on other users' computers. For example, a symbolic good can be a text (short or long), an image, an audio or a video. Text is the simplest form of a symbolic good: it can easily be created by the user through typing text on the keyboard, the transfer over the Internet is easy and cheap because of its small size, and it can be displayed on other users' computers easily. The difficulties of creating, transferring, and displaying other symbolic goods like image, audio, or video are higher but can nevertheless be managed.

An important point is that any of these media items (or a combination of several media items) can be the symbolic good which should be characterised by users. Agora can therefore be employed for many purposes –it is generic– and it is not restricted to one medium.

A symbolic good is called symbolic for two reasons: First, it does not represent a real, physical good and therefore symbolic is meant in contrast to physical. And second, it needs to be technically transferable over the Internet.

### 4.1.2 Characteristics of Symbolic Goods

Characteristics of symbolic goods can be expressed in various ways. In order to provide means to express as many characteristics as needed in applications run by Agora, a type system needs to be defined. Each characteristic that can be employed in Agora is of a certain characteristic type. This characteristic type consists of its name and the type of values that can be assigned. The value types can be number, number with relations, tuple, or string. These value types are described in the following.

Numerical values of type **number** are necessary as a standard for many applications and already cater for many characteristics. The kinds of data that can be represented using values of type number include the scales of measurement as defined by Stevens [193]:

- nominal, for unordered data (e.g., gender)
- ordinal, for totally ordered data (e.g., Likert scale)
- interval, for linear data with meaningful sizes of intervals (e.g., temperature or age)
- ratio, for linear data with a true zero point (e.g., proportions)

The **number with relations** type offers the possibility to manage numbers which have any kind of binary relations between each other. As an example, geographical regions can be represented as data of type number with relations: A unique number is specified as identifier for each region. Large geographical regions (e.g., countries) contain smaller regions (e.g., provinces) which themselves contain smaller regions (e.g., municipalities). The hierarchy of the regions can be represented as a binary relation, i.e., a set of tuples of a larger (parent) region and a contained smaller (child) region (e.g., (1, 2) specifies that region 1 has a child region 2). Employing this relation in Agora, an action on a parent region can result in a certain dividend being paid also if not exactly the same region is chosen but a smaller region within the same parent region. Also neighbouring regions can be considered similar for some actions, represented as another relation. The two relations hierarchy and neighbour can thus be represented as a set of two binary relations between two numbers –representing regions– each, which is a possible way for using number with relations.

Data of the type **tuple** offers the possibility to store, e.g., two-dimensional coordinates representing a geographical location or a 4-tuple representing two 2-dimensional coordinates of a rectangle in an image. Also a set of numbers can be stored as tuple.

The type **string** accommodates for textual data. In addition to numbers, strings are another basic type, offering much freedom for the implementation of applications built with Agora. Data that cannot be represented as numbers easily, such as textual descriptions, can thus also be handled in Agora. As an example for data of type string, tags can be characteristics of symbolic goods, representing a keyword description.

Summarising, the following defines the set of value types available in Agora:

**Definition 1** *The **set of value types** for characteristics is  $T$ , where a **value type**  $\tau \in T$  is any of:*

- *number, for any type of data representable as number*
- *number with relations, for any type of data representable as number for which one or more relations are defined*
- *tuple, for an ordered list of numbers*
- *string, for a sequence of characters*

In applications built with Agora, these value types can be employed for gathering data with characterisations or assessments. In order to do that, a set of characteristic types needs to be specified for the application. A characteristic type consists of a name of the characteristic, specifying what the content of the characteristic is, and a value type, specifying the type of data gathered:

**Definition 2** *Given a name of a characteristic  $n$  and its value type  $\tau \in T$ , a **characteristic type** is the tuple  $(n, \tau)$ .*

The set of characteristic types in specific applications may consist of one characteristic type for markets focusing on one specific characteristic, of several characteristic types for gathering different kinds of data, or of characteristic types that can be defined by users on demand. Each characteristic type must have a name unique within an application because the characteristic types are identified by their name.

For gathering data, a specific characteristic can now be defined consisting of a characteristic type and the value of the characteristic:

**Definition 3** *Given a characteristic type  $(n, \tau)$  and a value  $v$  of type  $\tau$ , the **characteristic** is  $c = (n, \tau, v)$ .*

As an example, a characteristic may be a geographical region, represented by the number 1, as  $c = (\text{region}, \text{number with relations}, 1)$ . Note that a characteristic is independent of the relations. For  $c$  itself it is irrelevant whether, e.g.,  $(1, 1)$  appears in a relation or not. Relations of the characteristic type number with relations can therefore be seen as a specific, data-independent view on the numbers.

To ensure the consistency of sets of characteristics, it should not be possible to have two characteristics with the same name  $n$  but a different value type  $\tau$  within the same set. If a set fulfils this property, it is called consistent:

**Definition 4** *A set  $C$  of characteristics is **consistent**, if*  

$$\forall c = (n, \tau, v) \in C \ \forall c' = (n', \tau', v') \in C: \text{ if } n = n' \text{ then } \tau = \tau'$$



For an application, the set of all existing characteristics should be consistent. Following the definition, a consistent set of characteristics can contain two characteristics with the same characteristic type and a different value. Therefore, it is possible to specify characteristics multiple times. For example, several geographical regions may be specified as characteristics for a symbolic good if it is common in multiple regions.

Finally, the following function extracts the characteristic types from a set of characteristics:

**Definition 5** *Given a set  $C$  of characteristics, let*  
 $characteristictypes(C) = \{(n, \tau) \mid \exists v : (n, \tau, v) \in C\}$

This function is provided for convenience and it is used in further definitions.

### 4.1.3 Characterisations, Assessments, and Further Actions

Based on the characteristics, actions define what is possible for users to do in applications built with Agora. Three actions most important for Agora are introduced as a basis: characterisations, assessments, and sales. Further actions can be defined depending on the actual application.

Characterisations –together with assessments– are the main concept of gathering meta-data of symbolic goods. A characterisation consists of a user and one or more characteristics of a symbolic good which the user specified. For example, in the context of metropolitalia, a characterisation represents the concept that a user specifies a certain geographical region in Italy in which a specific phrase is spoken. As an example of a characterisation with multiple characteristics, a user can specify that a certain phrase is mainly used by female, old, and educated people. The definition for a characterisation is thus the following:

**Definition 6** *Given a user  $u$ , a symbolic good  $g$ , and a consistent set of characteristics  $C$ , the **characterisation** is the tuple  $h = (u, g, C)$ .*

Characterisations can be employed for gathering any characteristics about symbolic goods. For the analysis and the aggregation of characterisations, it is important to determine whether two characterisations are compatible. The compatibility is given if the same characteristic types are used in the two characterisations, for example if each characterisation specifies a characteristic for the characteristic type (*region, number with relations*), and not more. In other words, characterisations with the same set of characteristic types are called compatible:

**Definition 7** *Given a symbolic good  $g$ , two characterisations  $h_1 = (u_1, g, C_1)$  and  $h_2 = (u_2, g, C_2)$  are **compatible** if*  
 $characteristictypes(C_1) = characteristictypes(C_2)$

Compatible characterisations can be combined to generate an aggregated characterisation of many users. All characterisations for a symbolic good together represent the market's view of the symbolic good.

An assessment is an enhanced form of a characterisation. Like characterisations, it consists of a user assessing one or more characteristics of a symbolic good, and additionally it includes an estimation which proportion of users are likely to assign the same characteristics as the user herself does. Through an assessment therefore an estimation of how many users agree to the user's characteristics is gathered, which conveys her perception of the given characteristics. Hence, assessments provide further meta-data on how widely known certain characteristics are in the opinion of users, which makes them interesting to be gathered and analysed.

**Definition 8** *Given a user  $u$ , a symbolic good  $g$ , a set of characteristics  $C$ , and an estimated agreement proportion  $p \in [0, 1]$ , an **assessment** is the tuple  $a = (u, g, C, p)$ .*

Similar to characterisations, assessments can be compatible to other assessments, and assessments and characterisations can be compatible. The definitions of compatibility within assessments and between assessments and characterisations resemble Definition 7:

**Definition 9** *Given a symbolic good  $g$ , two assessments  $a_1 = (u_1, g, C_1, p_1)$  and  $a_2 = (u_2, g, C_2, p_2)$  are **compatible** if  $\text{characteristictypes}(C_1) = \text{characteristictypes}(C_2)$*

**Definition 10** *Given a symbolic good  $g$ , an assessments  $a = (u_1, g, C_1, p_1)$  and a characterisation  $h = (u_2, g, C_2, p_2)$  are **compatible** if  $\text{characteristictypes}(C_1) = \text{characteristictypes}(C_2)$*

Two assessments may exist that are based on the same symbolic good, share the same characteristics and estimated agreement proportion, only the user is different. In this sense, assessments are not necessarily unique.

Assessments have a monetary value associated with them. The monetary value specifies how accurate the assessment is. If a user agrees with the gathered characteristics and assessments that are compatible to her assessment, the monetary value is high. The closer her estimation is to the proportion of users assigning the same characteristic, the higher the monetary value of the assessment. The computation of the monetary value is explained in [Section 4.2.1](#).

All data structures introduced so far are static, i. e., time-independent. Before introducing actions that change existing data, like the sale of an assessment, the state of an Agora application needs to be defined. The state consists of all data structures that exist at a certain point in time:

**Definition 11** *A **state** of an Agora application is the tuple  $(U, G, L, A, D)$  consisting of*

- a set of users  $U$ ,
- a set of symbolic goods  $G$  created by members of  $U$ ,
- a set of collections  $L$ ,

- a set of actions  $A$  created by members of  $U$  for members of  $G$  and  $L$ , and
- a set of dividends  $D$  distributed to members of  $U$  for members of  $A$  and  $L$ .

For convenience, the different types of actions are referred to as the following subsets of the set of actions  $A$ :

- $A_a \subset A$ : set of assessments
- $A_h \subset A$ : set of characterisations
- $A_p \subset A$ : set of purchases
- $A_s \subset A$ : set of sales

A state can be transformed into another state by actions. Two such actions that transform the state are sale and purchase. Assessments can be offered for sale for a user-defined price and purchased by other users. Thus users can create their own portfolio of assessments and gather assessments they deem to be important or valuable. For Agora, more meta-data about the consent to assessments are gathered through trades.

A trade is completed if one user offers one of her assessments for sale for a certain price and another user accepts this offer. Another possibility for a completed trade is that a user creates a purchase bid for another user's assessment of a certain price and the user owning the assessment accepts this bid. Both cases can be modelled using two tuples sale and purchase. An offer resp. a bid is represented by a sale resp. a purchase existing on its own. A completed trade is represented by both a sale and a purchase for the same assessment and with the same price, containing references to each other. The sale offers and purchase bids have no reference to each other.

**Definition 12** *Given a user  $u$ , an assessment  $a$ , and a price  $v$ , a **sale offer** is the tuple  $sale = (u, a, v, \emptyset)$ .*

**Definition 13** *Given a user  $u$ , an assessment  $a$ , and a price  $v$ , a **purchase bid** is the tuple  $purchase = (u, a, v, \emptyset)$ .*

Upon completion of a trade, the sale offers and purchase bids are transformed into a sale and a purchase by referencing each other, the ownership of the assessment  $a$  is transferred from the seller ( $u_s$ ) to the purchaser ( $u_p$ ), and the assessment is updated accordingly.

**Definition 14** *A state  $(U, G, L, A, D)$  which contains an assessment  $a = (u_s, g, C, p) \in A$ , a sale offer  $sale = (u_s, a, v, \emptyset) \in A$ , and a purchase bid  $purchase = (u_p, a, v, \emptyset) \in A$ , is transformed into a new state  $(U, G, L, A', D)$  with*

$$A' = (A \setminus \{a, sale, purchase\}) \cup \{a', sale', purchase'\}$$

where

- the assessment  $a' = (u_p, g, C, p)$ ,
- the sale  $sale' = (u_s, a', v, purchase')$ , and
- the purchase  $purchase' = (u_p, a', v, sale')$ .

Similar to this transformation of one state into another one, custom actions that modify the state can be defined accordingly for specific applications built with Agora. This offers the possibility to enhance the kinds of data gathered and to adapt Agora to the specific needs.

## 4.2 THE ROLE OF PLAY-MONEY IN AGORA

In Agora, some form of money needs to exist in order that assessments have a value, are comparable, and can be traded. Furthermore, money can act as an incentive for users. Agora therefore supports play-money, i.e. money that only exists within applications built with Agora and that does not represent real money issued by a government. This is intended to foster motivation similar to leaderboards in games. Users can gain play-money through assessments, which have a varying monetary value, through dividends, which are paid as rewards for certain user actions, and through selling assessments. The user's total amount of play-money is the sum of received dividends, prices of completed sales, and monetary values of own assessments minus the prices of completed purchases.

In the following sections, the monetary value of assessments in Agora, the payout of dividends for further user actions, and the trade with assessments are explained.

### 4.2.1 Monetary Value of Assessments

Assessments have a monetary value associated with them. The monetary value specifies how accurate an assessment is, i.e., how well the aggregated opinion of all users matches the estimated agreement proportion specified in the assessment. The monetary value of assessments is a substantial concept for evaluating single assessments. Assessments that represent the aggregated characterisations of all users for a certain symbolic good better, should be worth more money than others representing them worse. In the following, a schema for computing the monetary value of assessments is proposed that has been conceived to incorporate this important concept.

Before computing the monetary value of an assessment, the **agreement** of assessments needs to be defined. The agreement specifies to which degree all users currently agree to a user's assessment, i.e., specify the same (or similar) characteristics. If for example the assessment contains the characteristic  $c = (region, number\ with\ relations, 1)$  and all other users also specify characterisations with the same characteristic, the agreement value is 1, i.e., perfect agreement. If half of all users specify the same characteristic, the agreement

value is 0.5. And if no other users specify this characteristic, the agreement value is close to 0, i.e., no agreement. The agreement value can never be 0 because the assessment itself is included in the agreement value in order to yield a global view.

Different degrees of similarity between characterisations can be specified that introduce a more fuzzy approach of similarity. For example, a neighbouring region can be considered similar to a lesser degree than the region itself. This is catered for in the similarity function:

**Definition 15** *Given a state  $(U, G, L, A, D)$  where  $A_h \subset A$  is the set of characterisations, a **similarity function** is of the form*

$$s(h_1, h_2) : A_h \times A_h \rightarrow [0, 1]$$

The similarity function represents the similarity between the two characterisations  $h_1$  and  $h_2$ . As an assessment is a subtype of a characterisation, the similarity function can be applied to any combination of characterisation and assessment. The similarity function has to be adapted to the needs of the specific application. It can be the Kronecker delta function, which for a specified symbolic good would return 1 if the characterisations have the same characteristics and 0 otherwise. But also elaborate similarity functions can be defined depending on the context. Note that  $s$  is not necessarily commutative, i.e.,  $s(h_1, h_2)$  and  $s(h_2, h_1)$  can yield different results. Because of the dependence on the context, the similarity function cannot be defined here. The similarity functions for the applications on metropolitalia are for example defined in Sections 5.1.2 and 5.2.2.

Together, the agreement for an assessment –also called current agreement and calculated agreement in this thesis– is defined as follows.

**Definition 16** *Given a state  $(U, G, L, A, D)$  where  $A_a \subset A$  is the set of assessments and  $A_h \subset A$  is the set of characterisations, an assessment  $a = (u, g, C, p) \in A_a$ , and a similarity function  $s(h_1, h_2)$ , the **agreement** :  $A_a \rightarrow [0, 1]$  is defined as:*

$$\text{agreement}(a) = \frac{\sum_{h_i \in A_{h,g,C}} s(h_i, a)}{|A_{h,g,C}|}$$

where

- $A_{h,g,C} \subset A_h$  is the set of all characterisations that are compatible with  $a$  and
- $|\cdot|$  is the cardinality of the set.

Now, the calculated agreement can be compared to the estimated agreement proportion of the assessment to yield the **monetary value** of an assessment. The smaller the difference, the higher the monetary value. A simple, linear function for an assessment is the following.

**Definition 17** Given an assessment  $a = (u, g, C, p)$ , the linear function based monetary value of this assessment is:

$$value_{linear}(a) = 100 \cdot (1 - |agreement(a) - p|)$$

where  $|v|$  denotes the absolute value of  $v$ .

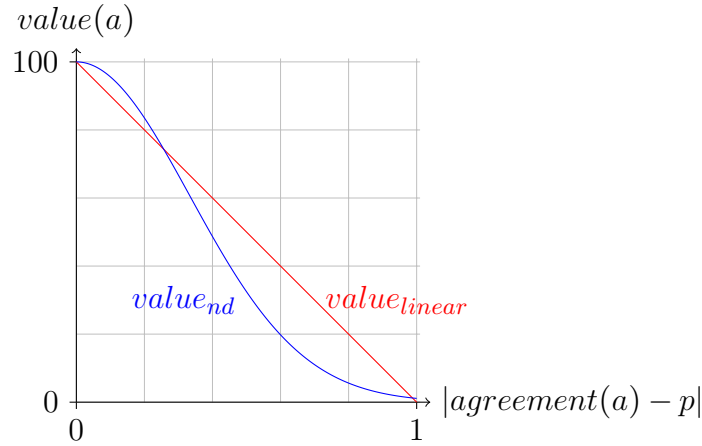
By multiplying a number like 100 to the pure difference, the value is more accessible to users than decimal numbers between 0 and 1. Also other functions to define the monetary value are possible, e.g., the density function of a normal distribution, which values close estimations higher and remote estimations lower than a linear function. It may thus promote good estimations even more.

**Definition 18** Given an assessment  $a = (u, g, C, p)$ , the normal distribution based monetary value of this assessment is:

$$value_{nd}(a) = 100 \cdot e^{-\frac{(agreement(a)-p)^2}{2\sigma^2}}$$

where  $\sigma^2$  is the variance of the normal distribution.

For example, if  $\sigma = \frac{1}{3}$ , the range of values is almost the same as in  $value_{linear}$ , only the distribution is different (as shown in Figure 4).<sup>1</sup> Note that the usual normalisation constant  $\frac{1}{\sqrt{2\pi\sigma^2}}$  is omitted because of the different value range.



**Figure 4:** Two functions  $value_{linear}$  and  $value_{nd}$  with  $\sigma = \frac{1}{3}$  expressing the monetary value of an assessment. The x-axis is the difference between the estimated agreement proportion and the calculated agreement.

The monetary value of assessments thus depends on the current agreement and therefore on the characterisations by all users. Because characterisations are provided by users as time goes by, the monetary value changes. If over time the agreement of an assessment diverges from the user's estimation, the user loses a part of the money the assessment was worth before. If it converges to her estimation, she gains money.

<sup>1</sup> To be exact, 99.73% of the values of  $value_{nd}$  are in the range. Thus if values are given as integers without decimal, the ranges of both value functions are the same.

### 4.2.2 Dividends

Dividends in Agora represent payouts of play-money to users. They are called dividends because, in a financial market, companies that are successful pay dividends to shareholders, and in Agora successful actions also lead to a payout. Such dividends are triggered through actions in Agora. One action can trigger several dividends being paid out to several users.

As an example, if a user generates a characterisation which resembles the current agreement for that symbolic good, she can –depending on the actual application– receive a dividend. Furthermore, users whose characterisation is validated through this user may also receive a dividend for their “successful” characterisation.

Dividends can therefore be seen as rewards for users for meeting the community’s collective opinion. Dividends must be seen separated from the monetary value of assessments. While the former are durable and do not change over time, the latter can (and probably will) change while other users characterise the same symbolic good.

Applications built with Agora must define which dividends are paid. For each action, a set of rules can be defined that control the payout of dividends for involved users.

### 4.2.3 Trading with Assessments

Assessments are a central part of Agora and Agora offers the possibility for users to trade them. Users can either trade assessments directly between each other or through a market maker in Agora.

For a **direct trade**, a user owning an assessment places a sale offer by setting a price for which it can be bought by other users. Then, others see that the assessment is for sale for the specified price. A user can accept the offer whereupon money and assessment are exchanged. The transaction is represented as Sale and Purchase in Agora’s data schema (see [Section 4.1.3](#)).

For purchasing an assessment, the user must have enough play-money to pay the specified price without getting a negative balance, i. e., having a negative amount of play-money. Naturally, the monetary value of a user’s own assessments cannot be used for purchasing other assessments. Thus, the user’s balance cannot get negative.

Direct trades are different to trades in which a **market maker** is involved. In financial markets, designated market makers typically act as agents who buy and sell shares in order to ensure the liquidity of the market, i. e., that other customers can buy and sell shares although no direct trade partner exists. Their task thus is “demand smoothing” [74]. In Agora, an automated market maker facilitates trading in cases in which users do not want to wait for a trade partner. In order to enable such trading, reasonable prices need to be computed for bids and offers.

One possibility is that the price paid to users offering an assessment must be lower than its monetary value and the price for offering an assessment must be higher than its monetary value:

$$\begin{aligned} \text{sale\_offer}(a) &= \text{value}(a) + \delta \\ \text{purchase\_bid}(a) &= \text{value}(a) - \delta \end{aligned}$$

where  $\delta$  is the premium the automated market maker demands for making profit itself, e.g.,  $\delta = 5$ . The premium ensures that users do not just buy (or sell) any assessment but only assessments they deem to be worth (or not worth), because in each case they lose money through the trade. This loss can, however, be amortised by the change in the monetary value of the assessment.

Another possibility is a forecast-based computation. In this case, the computed price is based on the trend of the monetary value of assessments. For example, the price paid to users may be higher than the assessment's monetary value if this monetary value has been rising during the latest characterisations.

On the one hand, the forecast-based computation seems to be more elaborate than the more simple computation and it may provide a better, more intelligent price. On the other hand, the task of analysing the trend of the monetary value of an assessment and estimating an appropriate price can and should be done by humans. Providing a forecast-based price to users could prevent them from thinking of an appropriate price for themselves. Furthermore, a good strategy consists in taking the users' prices and compare them to the prices computed by the forecast-based algorithm. The deficits and benefits of the algorithm can then be analysed and the algorithm improved. Applications built with Agora should therefore aim for the first, simpler strategy for the automated market maker. It furthermore is comprehensible for users and does not lead to confusion.

### 4.3 USER INTERACTIONS

Having introduced the core concepts and the role of play-money in Agora, now, the user interactions that are possible for applications based on Agora are described. Agora supports the following user interactions:

- creating symbolic goods
- characterising symbolic goods
- trading
- searching and browsing

Each user interaction focuses on gathering specific types of data. Some of these user interactions are related to actions as defined in the data schema in [Figure 3](#), namely creating symbolic goods (characterisation and assessment), characterising symbolic goods (characterisation and assessment), and trading



(sale and purchase), while searching and browsing is not yet related to these actions, but may well be through a custom action.

The four user interactions are described together with users' motivations and rewards in the successive sections.

#### 4.3.1 Creating a Symbolic Good

New user content is integrated in Agora by allowing users to create new symbolic goods. This is important to enliven the applications built with Agora so that they can grow both in the number of symbolic goods gathered and in the number of their users without being limited by the number of symbolic goods available for characterisation.

The specific method how users create symbolic goods depends on the type of the symbolic goods. For textual symbolic goods, a simple text field is sufficient. Audio or video needs to be recorded through a more sophisticated method, e.g., through the use of multimedia recording components like Flash [229] or HTML5's `getUserMedia` function (which is not yet available on all major Web browsers) [234, 258, 259]. The symbolic good is then transmitted to the server and stored for display or playback to other users.

In general, every registered user is allowed to create a new symbolic good. Protection against inappropriate or illegally distributed content that is prohibited by law or not suitable for redistribution is advisable. This can be achieved either by monitoring symbolic goods created and disabling them in such cases, or by including an option for users to report inappropriate content to the platform operators and responding to such reports by disabling symbolic goods. It depends on the prospected workload in terms of the amount of symbolic goods created which measure should be chosen for the application. If only few symbolic goods are created each day, monitoring can be an efficient measure. However, for very active platforms, the reporting of inappropriate symbolic goods by users should be preferred.

With a certain amount of symbolic goods, users may create symbolic goods that already exist in the system or only differ from existing symbolic goods in minor points. For example, minor differences can become apparent if the symbolic goods are phrases in language varieties for which the spelling is not standardised. The standard measure in Agora is to merge only symbolic goods that are identical, e.g., phrases that have the same spelling, as also minor differences in the symbolic good can make a big difference for it as a whole, e.g., a small spelling difference can make a big difference in word meaning. Merging means that if a user tries to create a redundant symbolic good and its characterisation, a characterisation for the existing symbolic good is created instead. This automatic merge is not displayed to the user because for her no difference exists. Other measures for identifying redundant symbolic goods can be incorporated into Agora if this is needed for specific applications.

In terms of users' motivation, stereotypes of persons can be differentiated having different motivations, the most prominent being *achievement*, *affiliation*, and *power* [70, 78, 199]. Achievement-motivated persons want to ex-

cel themselves or surpass others, affiliation-motivated persons want to receive recognition by others, and power-motivated persons engage in prestige-seeking behaviours for their own personal need for power [78]. A person usually includes these stereotypes in different emphasis, e.g., some persons are more power-motivated than affiliation-motivated while others are almost only achievement-motivated. The different user interactions in Agora address these stereotypes in different amounts.

Creating symbolic goods mainly caters for affiliation-motivated stereotypes. Users sharing own content get feedback from other users through their characterisations. The sharing mentality can furthermore be seen in the rise of platforms like Facebook [241] and Twitter [273], where users continually publish own content as well as links to content of others. Though these platforms differ from platforms built with Agora, sharing own symbolic goods is presumably motivated by similar reasons of affiliation: Users assume some symbolic good is interesting to other users, users are inspired by existing symbolic goods, users want to contribute to the platform, and users receive recognition by others through characterisations.

For creating a symbolic good, no direct reward is provided because a direct reward for its creation would encourage the creation of nonsense symbolic goods in addition to sensible ones. However, users can receive rewards through the assessment of created symbolic goods. After a user creates a symbolic good on an application built with Agora, she is prompted to characterise it and therefore specify meta-data. This way, during the creation of a symbolic good, a first assessment is created that forms a basis for further characterisations by other users. The prospect of the monetary value of the assessment provides a reward for the user to create it.

#### 4.3.2 Characterising a Symbolic Good

Characterisations and assessments of symbolic goods provide data and meta-data about symbolic goods and thus complement the creation of new symbolic goods. Only if both types of gathering data, the creation and the characterisation, are accomplished, platforms built with Agora can be successful.

For gathering assessments, users can specify characteristics of symbolic goods in a playful manner. Therefore, first a symbolic good is selected by Agora – randomly or by an algorithm that chooses well suited symbolic goods – and displayed to the user. Such an algorithm can for example prefer symbolic goods having only few assessments so far or select specific symbolic goods for quality control. Randomness is important for ensuring the quality of gathered data. Further details on such quality control measures are given in [Section 4.5](#).

When the symbolic good is displayed, the user is prompted to specify one or more characteristics. Which characteristics, in which order, and whether sequentially or in parallel needs to be defined depending on Agora’s use case. Two exemplary playful applications with a certain set of characteristics in the context of metropolitalia are introduced in [Chapter 5](#). The kind of input control –for example, a text field, an option list, or a geographical map with regions–

can be specified as well. In addition to predefined input controls, custom ones can be implemented for tailoring the application to its specific needs. Besides specifying characteristics, users can estimate which proportion of users choose the same characteristics as they do, and therefore create assessments.

After the user specifies characteristics and, optionally, the estimated agreement proportion, feedback on the user's assessment can be provided. Such feedback can comprise the current status of the characterisations by all users as collective value, a message whether the user has been specifying the characteristics according to the community, and the display of monetary or non-monetary rewards. Non-monetary rewards can for example be badges that are awarded to users for a certain achievement. For the user's characterisations, dividends may be paid as a reward for her, depending on the application's rules. Additionally, assessments provide the possibility for gaining more play-money.

Several symbolic goods can be assessed in a row in several rounds, and a summary can be shown at the end of all assessments. This method leads to a higher response rate compared to specifying single assessments. If specified for the application, users are able to skip symbolic goods they do not want to assess, and time limits can be enforced.

Motivating users for characterising symbolic goods is crucial for gathering much and manifold data. Mainly achievement-motivated stereotypes are addressed with characterisations: Monetary or non-monetary rewards can motivate users to provide characterisations and to continue characterising symbolic goods until they reach a certain position in the list of the best users. Moreover, users get access to symbolic goods which might be unknown to them and enjoyable. This is the case for symbolic goods that are important for the specific user. Also affiliation-motivated stereotypes are addressed: feedback provided can motivate users to contribute data because on the one hand their characterisations are of interest and evaluated and on the other hand they learn how other users characterise a symbolic good and thus are recognised for their own characterisations.

It can be specified for each application which rewards are provided and what value they have through dividends. For assessments, a schema for calculating their monetary value has been introduced in [Section 4.2.1](#). This schema needs to be adapted to the situation in terms of the similarity function and parameters. Then, monetary rewards can directly be displayed as feedback to users.

### 4.3.3 Trading

Trading of assessments can be initiated both by potential sellers and buyers. Sellers can offer an assessment for sale by selecting one of their own assessments and entering a price they want to achieve. This offer is displayed to other users of Agora's platform and can be accepted by them. Besides, Agora's market maker can purchase the assessment and offer it for sale (see [Section 4.2.3](#)).

Buyers can place a bid on any assessment they stumble upon on Agora platforms. If the assessment is already up for sale, they can accept the sale

offer or place a bid with a different price. The potential trade partner is informed of the bid and can accept or decline the bid. She furthermore can offer the assessment for sale at a different price, whereupon the potential buyer is informed.

The main motivation for trading with assessments is power. First, one goal on platforms built with Agora is to maximise wealth and have a high position in the list of the best users. Trading offers the possibility to gain more money by selling assessments with a negative forecast (in the user's opinion) and buying those with a positive forecast. Users are rewarded for high quality assessments and the speculation with them. Furthermore, having a portfolio of assessments currently being worth much, i. e., assessments of "high quality" regarding the community's opinion, can be a motivation for selling assessments that are not worth much and buying assessments that are worth much. Another incentive can consist in holding assessments that are meaningful to a user or have a certain symbolic value to her. This might also motivate affiliation-motivated stereotypes. For example, users speaking a certain dialect might accumulate assessments representing this dialect and trade for this goal. Public display of own assessments can further enhance the gathering of assessments meaningful to users. This can be done by grouping assessments in collections and enabling public visibility for some collections.

#### 4.3.4 Searching and Browsing

Another user interaction consists in searching and browsing through existing symbolic goods and their characterisations and assessments. Platforms built with Agora have the potential to gather many symbolic goods together with characteristics of many people. Searching and browsing functionalities help to unveil the gathered data and meta-data to the public.

In detail, a search must be possible for symbolic goods, for characteristics, and for combinations thereof. For symbolic goods in textual form, a reasonable way is to allow for full-text search in the text. For symbolic goods in audio or video form, voice or text recognition algorithms can be applied in addition to letting users provide a textual description. The characteristics are searchable by providing the value of the characteristic, e. g., "female" and "male" for a gender characterisation or the name of a region for a geographical characterisation. Ranking search results in the best order possible can be difficult. The users' assessments help to generate a better ranking than that achievable without the extensive meta-data. For example, when searching for "female", symbolic goods assessed as more female –i. e., the ratio of the number of users choosing the characteristic "female" to the number of users choosing the characteristic "male" is higher– than other symbolic goods can be ranked higher. The search results should then be displayed in a user-friendly way that enables easy navigation through the results.

The navigation through the results can be further enhanced by linking symbolic goods that are similar in some way, for example that have words in common, share certain characteristics, are in the assessments list of the same

user, or are in one collection. Then, users can navigate directly from one symbolic good to other related ones. Such strategies of finding relevant objects are called browsing [9, 80, 81]. Other browsing techniques can be incorporated, for example for browsing by characteristics, e.g., a geographical map can be displayed and regions highlighted in which symbolic goods have been characterised. A specific type of browsing can be achieved using recommender systems, techniques that provide suggestions for items that are deemed to be useful for the user [170]. They have become well known through recommendations in online stores, foremost Amazon, for similar items, as “users who bought x also bought y”. Employing recommendation techniques for similar symbolic goods can enhance the value of browsing through symbolic goods for users.

Agora is designed to cater for both, searching and browsing, and provides means to integrate these user interactions into the actual applications.

## 4.4 AGORA'S SIMILARITY TO A FINANCIAL MARKET

Agora is similar to a financial market like Wall Street in two ways, in the similarity of assessments to derivatives and in the existence of trading.

To understand the similarity to derivatives, the functioning of derivatives in a financial market [23] is briefly explained. In a financial market, participants can trade financial securities such as shares, bonds, or derivatives. Participants for example speculate on the valuation of companies, on the value of goods like gold or oil, and on exchange rates of foreign currencies. The prices of securities in a financial market depend on the supply and the demand of the security. Derivatives, specifically, have no value in themselves (like company shares have) but *derive* their value from other entities under conditions specified in contracts between two trade partners [91]. One kind of derivatives are options in which buyers are entitled (but not obliged) to buy or sell a certain quantity of an underlying security at a specified strike price on or before a specified expiration date. For this right they pay a premium to the seller. Options are similar to assessments in Agora, with which market participants speculate on the community's opinion of characteristics of symbolic goods. The estimated agreement proportion of an assessment resembles the strike price of an option and its symbolic good resembles the underlying security. For assessments, no sale or purchase of the underlying symbolic good is intended and no expiration date exist. The intention to reach a certain price for an assessment / an option, however, is similar because upon reaching the specified agreement proportion in Agora the user gains the most money from her assessment. Furthermore, the actual price for which an option is traded on the market varies with the current value of the underlying security. The same exists in Agora: the monetary value of assessments varies while users assess the underlying symbolic goods and therefore influence the actual agreement.

Trading is the second similarity between Agora and a financial market. Assessments can be traded by Agora's users and the prices can be chosen by users. The monetary value of assessments suggests a reasonable price for trades, though this price is not binding. Buyers can, for example, indicate that an assessment is underrated by offering a price higher than its monetary value. Such differences can give useful insights into changing characteristics of symbolic goods as well as opinions of certain Agora users.

Due to Agora's similarities with financial markets, many users are already familiar with the market models in Agora. And for users unfamiliar with financial markets, the user interface eases their start with the market mechanisms. While Agora is similar to a financial market in several ways, Agora also differs from a financial market in the following points: Agora is a play-market, that is, no real money is involved. Furthermore, assessments do not need to be purchased but can also be obtained by users through engagement on the platform. As effort, the user has to characterise symbolic goods instead of spending money.

## 4.5 ENSURING THE QUALITY OF THE GATHERED DATA

Data quality is of high importance in crowdsourcing applications for data gathering, besides the quantity of the data. In general, good quality data in crowdsourcing are data that are reliable and correct. In crowdsourcing applications not all gathered data is of a reasonable quality, because participants may not understand the application and its functioning, they may just try it without sincere motivation, they may be motivated but ignorant or not smart enough, they may cheat, or they may deliberately give wrong answers, either to gain some advantage or to sabotage the system. Therefore, quality control measures are often necessary and employed in crowdsourcing applications.

Lease [126] reasons about such quality control measures from a machine learning perspective and identifies five areas of concern: application developers should be aware of human factors, include automated quality control and cheat detection (though carefully), communicate annotation guidelines clearly, optimise worker and task organisation, and take care of the minority voice without disregarding it as noise.

In Agora, measures incorporating these findings are included. One general problem for data gathered on crowdsourcing platforms for field research is that quality is difficult to assess because the characteristics for a symbolic good cannot be verified without doing field research itself. For example, no algorithm exists that reliably assigns geographical regions to dialect expressions. If existing dialects change the reality might be different again. Therefore, platform operators must in general rely on the contributors' motivation to enter quality assessments. Additionally, sensible incentives to cater for human factors (see [Section 4.3](#)), automated strategies to prevent cheating, applying gold stan-

dard measures, and filtering data after gathering it help to obtain quality data. Strategies to prevent cheating and gold standard assessments are incorporated into Agora, data filtering can be enabled if needed. These three measures are described in the following sections.

#### 4.5.1 Strategies to Prevent Cheating

Cheating is malicious behaviour with the goal to obtain advantages of some kind in a manner not intended by the system operators. Few strategies to prevent cheating are included in Agora because every anti cheating measure has some drawback and most measures can be circumvented by users in an elaborate or time-consuming way. The best option to prevent cheating is to not provide incentives that could attract such behaviour. If for example real money as reward was promised, the criminal energy to use the platform fraudulently is spurred. Therefore, platforms built with Agora should refrain from rewarding users with real money or other monetary rewards. Focusing on the correct incentives instead helps to gather quality data.

Nevertheless, some anti cheating measures must be taken in Agora. First, symbolic goods cannot be selected by the users' choice for characterisation. Instead, they are selected by Agora randomly or pseudo-randomly. If users were able to choose a symbolic good they want to assess, they could look at the gathered characteristics and then deliberately choose characteristics and an estimated agreement proportion such that the monetary value of their assessment is maximised. Therefore, symbolic goods are selected by an application-dependent algorithm in Agora. Symbolic goods can be selected purely at random or pseudo-randomly with a prioritisation of symbolic goods that fulfil certain criteria. For example, "gold standard" symbolic goods that can be used to check the quality of the user can be pushed to users once in a while (further details in the next section).

Furthermore, a user cannot assess the same symbolic good more than once. Though the probability for users to be shown the same symbolic good twice is quite low with a purely random selection, the possibility has to be eliminated as far as possible. Therefore, the selection algorithm needs to take care of not choosing the same symbolic good more than once as well, e. g., by keeping history based on the user account. This is the way it is done in Agora. It should be noted that it is technically not possible to exclude the possibility that an anonymous user connects to the platform via several computers or different Web browsers without logging in. In this case, users might get the same symbolic good twice, but even then with a very low probability.

#### 4.5.2 Gold Standard Assessments

As quality control measure for paid crowdsourcing, Oleson, Sorokin, Laughlin, Hester, Le and Biewald [155] suggest to use so-called gold standard questions for which the answers are known and furthermore to compute the set of gold



standard questions during evaluation automatically from a small seed gold set. Specifically, the workers of the underlying CrowdFlower platform[237] must first correctly answer a certain amount of these questions as training before they are allowed to answer real questions. In between the real questions some gold questions are asked in order to check the worker’s accuracy and give feedback to the worker if she is wrong. Questions with a high agreement are turned into gold questions. Their experimental evaluation shows that this method works and is a scalable way to achieve quality.

A similar approach is incorporated into Agora. Assessments of symbolic goods that have been characterised identically (or very similarly) by several users automatically become gold standard assessments. These symbolic goods are displayed to users every few rounds and their characterisation is compared to the community’s assessment. Users who do not choose the “correct” characterisation are not hindered to go on engaging on the platform –like it is done during the training period of the CrowdFlower mechanism– but the result is instead saved for later filtering of the data. Also in Agora, feedback is provided to users about their assessment, similar to CrowdFlower. One difference is that users do not know that they are gold standard assessments, while in CrowdFlower such an information is given.

Additionally to being a quality control measure, Agora’s gold standard approach increases the users’ motivation for continuing to engage on the platform for the following reason. For gold standard assessments, it is guaranteed that a certain amount of characterisations exist and therefore that the symbolic good is reasonably well characterised. Users get an elaborate feedback for these symbolic goods, based on the existing characterisations of the other users. If users characterise other symbolic goods that have been added to the platform only recently and thus have few assessments, an elaborate feedback on the symbolic good’s characteristics is not possible and users might get bored if only such symbolic goods were displayed. Therefore, the gold standard approach provides diversity in terms of displayed symbolic goods to users.

### 4.5.3 Filtering Data

The quality control measure of employing gold standard assessments is one basis upon which to filter data. Another way is to discard data from users who did not complete the characterisation of several symbolic goods but stopped after one or a few. The probability is high that such users only experiment with the platform without entering valuable data. The threshold must be low enough in order to keep all valuable data.

Moreover, the users’ trustworthiness can be analysed. Data from users who created (reasonable) symbolic goods on the platform can be considered trustworthy because the creation of symbolic goods takes more effort than the characterisation of existing symbolic goods. Users can also be considered to be trustworthy if they pass the gold standard tests or if they provided assessments having a high monetary value. Filtering data by untrustworthy users can furthermore enhance the data quality.



Another basis on which to filter data can be the location from where the user participating on the platform accesses the Internet. To gather this data, the location of each user's IP address is looked up in a geolocation database [257]. If for example a certain language is under research, data may be filtered to only contain data by users from IP addresses in a region where this language is spoken. Such filtering will probably discard data by valuable users from outside this region, especially because of the increasing amount of globalisation users access the Internet from everywhere. However, in some situations it might be useful to know where the users are from and base the filtering measures on this data. For these reasons, this measure is not active in Agora, can be enabled, and should be used with care.

These filtering techniques can be employed to discard data irrelevant to the research question that has been stated. The exact selection of techniques for an evaluation has to be chosen carefully by the researcher in order to yield valuable results.



## 5

MERCATO LINGUISTICO AND  
POKER PAROLE – TWO  
COMPLEMENTARY  
CROWDSOURCING  
APPLICATIONS BUILT WITH  
AGORA

Aspects of the content of this chapter have been published in [24, 107, 109–111]: [Section 5.1](#) is based on [24, 107, 109–111], [Section 5.2](#) and [Section 5.3](#) are based on [24, 110], [Section 5.4](#) is partly based on [107], [Section 5.5](#) is derived from [110], and [Section 5.6](#) extends parts of [24, 110].

On the platform metropolitalia, Agora is used for running two crowdsourcing applications, Mercato Linguistico and Poker Parole. In both, the symbolic goods under research are phrases –that is, sentences or parts of sentences– in Italian language varieties. The applications require written sentences but an extension with spoken sentences is possible. This extension would not require changes in the logic but only additional user interfaces for collecting and rendering spoken language. The usage of the word “play” is, in the following, limited to obviously playful actions [175] in the applications in the following.

Mercato Linguistico and Poker Parole are introduced in the following sections and the kinds of data gathered and user incentives are explained.

## 5.1 MERCATO LINGUISTICO

Agora is first used as operating system of an application called *Mercato Linguistico*, Italian for “linguistic market”. The market works in the following way: the better phrases and their characteristics are recognised by the user community, the more successful is a user expressing the same belief. Thus, mainly phrases with widely acknowledged linguistic traits are gathered with Mercato Linguistico. The goals of Mercato Linguistico are to gather new phrases and to encourage users to share their own assessments on new or existing phrases. Specifically, the user is asked to indicate where a phrase is spoken, how many people recognise the phrase as being from that location, which word(s) of the phrase are linguistically distinct, and who the speakers are in terms of age, gender, and level of education. A screenshot of Mercato Linguistico during a geographical characterisation of a phrase is shown in [Figure 5](#).

Mercato Linguistico has been under development since the first public version appeared in August 2012. During the first year, the acceptance of the features



**Figure 5:** Mercato Linguistico during the choice of a region for the displayed sentence. The currently selected region (northern Italy) is highlighted in blue.

of the first version was tested. In this version, users had the possibility to adjust the estimated agreement proportion of assessments once a day, based on the feedback from the market. This was implemented with the intention that the users could correct their estimations and are rewarded for visiting the platform again. As this feature has been used by very few users, it was removed from Mercato Linguistico. Instead, trading has been activated in November 2013, making it possible for users to sell and purchase assessments and thus to create own collections of assessments.

### 5.1.1 User Interactions

Mercato Linguistico offers several ways for users to interact with the platform. Four distinct web pages exist that offer all user actions defined in Agora. The individual web pages are for

- adding new phrases,
- characterising existing phrases,
- reviewing own assessments and selling them, and
- purchasing assessments which are for sale.

The user interface for **adding new phrases** is displayed in Figure 6. Two input fields for the phrase itself and the translation into standard Italian are provided. Furthermore, the geographical region can be selected graphically and

COME GIOCARE CHI SIAMO AREA COMMENTI REGISTRATI ACCEDI

0 PUNTI  
LIVELLO: SPETTATORE  
Riconosci un luogo per diventare principiante.

**metropol**  
**italia**  
social language tagging

Aggiungi un'espressione che ritieni tipica di un luogo o di una determinata regione.

Scommessa: Quale sarà la percentuale dei giocatori che conosce l'origine della tua espressione?

Si dice in:  
> Nord

OK Indietro

Per una scelta più precisa clicca nuovamente sulla cartina oppure conferma.

Almanco finisci di mangiare.

L'espressione significa: Almeno finisci di mangiare.

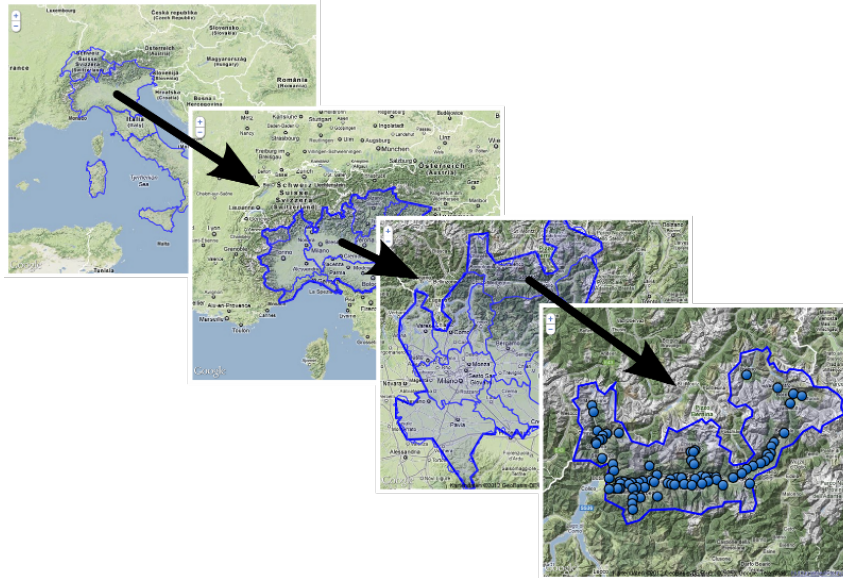
Invia la tua espressione

**Figure 6:** User interface for adding new phrases to the metropolitalia platform. The dialect phrase “Almanco finisci di mangiare” (English: At least finish eating) is entered with its translation into standard Italian (the only dialect word is “almanco”), the geographical region selected (northern Italy), and the estimated agreement proportion (80%) chosen.

the agreement proportion can be estimated. Upon submitting these data the assessment is added to the user’s list of assessments and to the phrases that can be characterised by other users.

The **characterisation of existing phrases** is the second way of interaction and the interaction with the most playful character. Mercato Linguistico is played in three rounds, which turned out to be a reasonable number for casual players who want to see the results quickly as well as for engaged players who continue playing with another series of rounds after getting feedback. Each round, one phrase is presented to the user which she has to assess. The following can be done by the user one after the other:

- choosing the geographical region where the phrase is spoken (see [Figure 5](#)),
- specifying her belief what percentage of users assigns the same region (the estimated agreement proportion),
- selecting individual words of the phrase that guided the user’s geographical mapping, and



**Figure 7:** Stepwise focusing on smaller regions: broad geographical regions, political regions, provinces, and municipalities.

- characterising social attributes of speakers of the phrase (age, gender, level of education).

For choosing a region, the user interface provides a top-down approach – stepwise focusing on smaller regions: broad geographical regions, political regions, provinces, and municipalities (as displayed in [Figure 7](#))– as well as a bottom-up approach –an input field with automatic suggestions of regions after entering the first few letters of a region. Some region names are not unique over all hierarchies, e.g., “Roma” is a province as well as a municipality. To avoid an ambiguous choice, the automatic completion of the input field displays the name of the region as well as the type of the region (i.e., broad geographical region, political region, province, or municipality).

Each user action is optional, i.e., can be skipped, to give the user freedom in her choice and to prevent false data if a user does not know what to choose. After all rounds, a summary is shown in which the user can see how other users characterised the phrases.

The user interface for **reviewing own assessments** is displayed in [Figure 8](#). The user’s assessments are displayed together with their monetary value, the current agreement, and furthermore the individual characterisations of single players. Moreover, the sale of assessments is possible here. The user has the choice either to offer an assessment for sale for an adjustable price or to sell it directly to the automated market maker for a fixed price.

[Figure 9](#) displays the user interface for **sale offers**. Here, the user can browse through assessments that currently are for sale, investigate their current agreement including details about the individual characterisations, and purchase assessments if intended.





**Figure 8:** User interface for reviewing own assessments. The user's assessment of the selected phrase "Quella ragazza è una pittima" (English: This girl keeps complaining about nothing) is worth 88, which is close to the maximum of 100. Her estimation of 32% agreement is nearly met, as 37% of metropolitalia's participants agree that the phrase is from southern Italy. The phrase can be offered for sale for an adjustable price and it can be sold directly for 83.

### 5.1.2 Customisation of Agora's Generic Aspects

Mercato Linguistico is built with Agora and therefore the generic aspects of Agora are customised to fit Mercato Linguistico's actual goals. The customisations of symbolic goods, their characteristics, actions and dividends, and the monetary value of assessments are explained in the following sections.

#### 5.1.2.1 Symbolic Goods

The symbolic goods under survey in Mercato Linguistico are phrases in Italian language varieties. The varieties can be dialects and other language varieties (see Section 3.3). "Phrase" is used in the sense of "symbolic good" in this



**Figure 9:** User interface for purchasing assessments. The selected assessment could be purchased for a price of 100.

chapter, being the customised form of Agora's generic symbolic good. For each phrase, its corresponding standard Italian form is available. Users who do not understand a dialect phrase thus can reveal its meaning in standard Italian.

#### 5.1.2.2 Characteristics

The following five characteristics are relevant in Mercato Linguistico.

- The *geographical region* is the most important characteristic, describing in which geographical region or city phrases are used. The value type  $\tau$  is *number with relations*. Each region is represented by a number. The broad geographical regions Switzerland, northern Italy, middle Italy, southern Italy, Sardinia, and Sicily represent the highest level regions where varieties of the Italian language are spoken. These broad linguistic regions include administrative regions which furthermore include provinces and municipalities. This division is an approximation of the division by language varieties. In the case of Italy the administrative approximation corresponds closely to linguistic boundaries and is thus a natural



choice. Furthermore, most Italians are familiar with administrative regions, provinces and municipalities, and a different division could confuse users. The hierarchy is visualised in [Figure 7](#). Besides the hierarchical relation, a neighbour relation between regions that have a common border exists.

- The *set of relevant words* specifies the individual words of a phrase that guided the user’s geographical mapping. The value type  $\tau$  is *tuple*. The numbers stored in the tuple represent the positions of the relevant words in the phrase.
- The *age* of speakers using the phrase can be characterised as young, old, everyone, and unknown. These values are stored as generic value type *number*, encoded as young=1, old=2, everyone=3, and unknown=0.
- The *gender* of speakers using the phrase can be characterised as female, male, both, and unknown. These values are stored as value type *number* as well, encoded as female=1, male=2, both=3, and unknown=0.
- The *level of education* of speakers using the phrase can be characterised as uneducated, educated, everyone, and unknown. These values are stored as value type *number* as well, encoded as uneducated=1, educated=2, everyone=3, and unknown=0.

For each social attribute, age, gender, and level of education, there are four possible values: the two extreme values (e.g., young and old), the value unspecific (e.g., speakers of every age), and the value unknown. This limited amount of values has been suggested by a Romance linguist in personal communication. He reasons that these values are sensible for evaluation, they provide clear instructions to user, and the user is not overwhelmed by too many options. Although the user’s view of, e.g., “old” might depend on her own age, the value still provides valuable feedback when aggregated over many users.

Formally, the set of characteristic types (according to the definition in [Section 4.1.2](#)) is:

$$\{(\textit{geographical region}, \textit{number with relations}), \\ (\textit{set of relevant words}, \textit{tuple}), \\ (\textit{age}, \textit{number}), \\ (\textit{gender}, \textit{number}), \\ (\textit{level of education}, \textit{number})\}$$

### 5.1.2.3 Actions and their Dividends

In Mercato Linguistico, the following actions are defined as part of Agora’s generic schema (ordered by the appearance during a game round):

- Unveiling of meaning: A user reveals the standard Italian form corresponding to a phrase. This standard Italian form is shown upon clicking

the question mark next to the phrase (see [Figure 5](#)). By including this unveiling of meaning as an action in the data schema, this data is saved for later usage. It could for example make a difference for an evaluation concerning the prominence of phrases. No dividend is associated with this action and it is not shown to the user that the action is recorded.

- **Geographical characterisation:** A user selects a geographical region where she deems a phrase is used. The region is saved as a characterisation and the characterisation is evaluated. A dividend is paid to the user who characterises the phrase, depending on the agreement for her characterisation. The agreement is calculated based on a fuzzy similarity which takes neighbouring and descendant regions into account (see next section). If the agreement is at least 50%, 10 points are paid, if the agreement is at least 30%, 5 points are paid, otherwise none. This dividend provides direct feedback for the user on how well her characterisation matches with the community opinion. Direct feedback is not available for new phrases.
- **Geographical assessment:** After a geographical characterisation, a user specifies the estimated agreement proportion, i.e., which percentage of users choose the same geographical region as she does. The creation of an assessment leads to its inclusion in the list of the user's own assessments. Its monetary value is calculated (see next section) and the resulting value displayed in the user's current amount of play-money. No dividend is paid as part of creating the assessment.
- **Word selection:** A user selects individual words of a phrase which guided the user's geographical characterisation and thus are linguistically relevant. A dividend is paid to the user, depending on how many users select the same words as being relevant. If at least 50% of the users select the same words, 10 points are paid, if at least 30%, 5 points are paid, otherwise none. Also here, the dividend provides feedback on the user's match with the community opinion.
- **Social characterisation:** A user selects age, gender, and / or level of education of speakers using the phrase. The dividend system here is analogous to the one for the word selection. The user is rewarded for specifying the same characteristics as the community. For each social attribute, 10, 5, or 0 points are awarded, for at least 50%, 30%, or 0% match with the community opinion. In the current version of Mercato Linguistico social assessments are not integrated. However, such an extension is conceivable and can be achieved with little effort.
- **Sale:** A user offers an assessment for sale for a user-defined price or directly sells it to the automated market maker for a price that is equal to the current agreement minus 5 (i.e.,  $\delta = 5$  in the equations in [Section 4.2.3](#)). The creation of a sale offer leads to the assessment being shown to other users for purchase. A direct sale leads to an immediate exchange of play-money and assessment with the automated market

maker (i. e., the platform itself). The assessment is then offered for sale by the market maker with a price premium of 5. No dividends are involved in a sale.

- Purchase: A user purchases an assessment that is offered for sale for the given price. Play-money and assessment are exchanged with the user offering the assessment (or with the market maker) and the sale offer is converted to a sale. No dividends are involved in a purchase.

#### 5.1.2.4 Monetary Value of Assessments

The general method for computing the monetary value of assessments is described in [Section 4.2.1](#). It remains to define the function  $s(a_1, a_2)$ , which represents the similarity between two assessments, and to choose proper parameters for the  $value_{nd}$  function.

The similarity function for geographical assessments in Mercato Linguistico represents a fuzzy similarity in which not only the same geographical locations are similar but in addition neighbouring and descendant regions. The similarity function for assessments on geographical regions is defined in Mercato Linguistico as:

$$s(a_1, a_2) = \begin{cases} 1.0 & \text{if } c_1 = c_2 \\ 0.8 & \text{if } (c_1, c_2) \in R_{neighbour} \\ 0.8 & \text{if } (c_1, c_2) \in R_{descendant} \\ 0.5 & \text{if } (c_2, c_1) \in R_{descendant} \\ 0.6 & \text{if } \exists c \in C_g : \\ & (c_1, c) \in R_{neighbour} \wedge (c, c_2) \in R_{descendant} \\ 0.3 & \text{if } \exists c \in C_g : \\ & (c_2, c) \in R_{neighbour} \wedge (c, c_1) \in R_{descendant} \\ 0 & \text{otherwise} \end{cases}$$

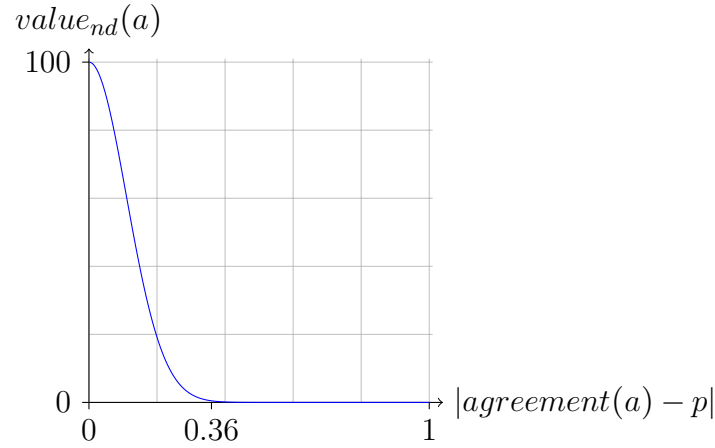
where

- $c_i$  is the geographical characteristic in assessment  $a_i$ ,
- $C_g$  is the set of all possible geographical characteristics, and
- $R_{neighbour}$  resp.  $R_{descendant}$  is the neighbour resp. descendant relation for geographical characteristics.

The similarity values for the individual cases have been chosen to convey the notion of similarity, i. e., geographical regions that are the same have full similarity (1.0), neighbouring regions are still very similar (0.8), parent or child regions are partly similar (0.8 resp. 0.5), combinations of neighbouring and child relationships a bit similar (0.6 resp. 0.3), and regions that have no close relation are not similar (0). The given order of the cases by their descending value is important for the similarity measure while the individual values may be adjusted to slightly different values.

Note that this similarity-like function is not symmetric, e. g., if a user characterises a phrase as being from a smaller, descendant region than another user, the similarity for her is 0.8, whereas the similarity is 0.5 if she characterises the phrase as being from a larger, ancestor region than another user. The difference in similarity reflects the fact that choosing a larger region is easier than a smaller region and that users are rewarded if they can characterise a phrase more precisely than others.

The monetary value of a user’s assessment is calculated if at least one other user has given a characterisation for the same symbolic good, otherwise the monetary value is 0. For computing the monetary value based on the agreement, the  $value_{nd}$  function is used with a value of  $\sigma = 0.11$  (see Figure 10). This value of  $\sigma$  results in a normal distribution that has the value 100 at no difference between the calculated agreement and the estimated agreement proportion, that reaches the value 1 at a difference of 0.36, and that has the value 0 at a difference  $> 0.36$ . This can be seen as an appropriate deviation for which to not award users money anymore.



**Figure 10:** The function  $value_{nd}$  with  $\sigma = 0.11$  expressing the monetary value of an assessment. The x-axis is the difference of the estimated agreement proportion and the calculated agreement.

### 5.1.3 Success on Mercato Linguistico: Keynes’ Beauty Contest

For being successful on Mercato Linguistico, one has to submit phrases with characteristics that many other users of Mercato Linguistico are likely to agree with. As a consequence, success on Mercato Linguistico depends on how one is skilled at forecasting others’ conceptions. This is a typical case of a “beauty contest”, as this experiment has been called by Keynes. In a beauty contest, as it has been performed in U.S. newspapers in former times, only those participants entered for a lottery drawing that chose the person that most other participants chose as being the most beautiful [102]. More generally put, in a speculative market participants reflect on each others’ behaviour and adapt

their behaviour accordingly. This leads to participants not necessarily revealing their own opinions but instead their estimations what the crowd of all participants would estimate [102].

However, while the beauty contest analogy was meant by Keynes as a criticism of speculation on financial markets, a beauty contest-like speculation contributes to the aim of Mercato Linguistico. Indeed, in linguistic field research the perception of the community's opinion provides further important information in addition to the opinion of a single speaker. In other words, for linguistic field research speculating users are welcome.

## 5.2 POKER PAROLE

A second application called *Poker Parole*, Italian for “word poker”, is also based on Agora, gathering complementary data.

Poker Parole shares many properties of its concepts with Mercato Linguistico, with one exception: While success on Mercato Linguistico comes from submitting commonly recognised phrases, on Poker Parole it comes from submitting phrases that most users are not likely to properly localise. Such phrases are equally important for linguistic research because they represent a minority opinion and therefore need to be gathered as well. The two applications therefore complement each other in the data they gather.

Poker Parole has been published in April 2013, after Mercato Linguistico had attracted its first users and yielded first data.

### 5.2.1 User Interactions

The set of three distinct web pages is similar to that of Mercato Linguistico. The user interactions possible within each one are different and no trading of assessments is allowed in Poker Parole. Trading is excluded in this version in order to provide a simple application to the users.

For **adding new phrases**, the user is asked to give a phrase with a characterisation that is hardly known by anybody living outside the chosen geographical or social regions. So the speculation consists in telling the community: “I guess that most of you won’t be capable of correctly recognising the characteristics of the following sentence.” By giving a phrase with a characterisation, an assessment is created with an estimated agreement proportion of 0. Apart from the instructions and that the estimated agreement proportion cannot be specified, the interface is the same as in Mercato Linguistico.

The **characterisation of existing phrases** is also played in three rounds. Each round, one phrase is presented to the user which the user has to assess. The user can either skip the phrase as being unknown to her or choose a geographical region where the phrase is spoken according to her opinion. Upon choosing a region, she receives feedback about her choice. If she chose the same region as the creator of the assessment, she is rewarded with money: 10 for

assessments that have been revealed by many other players, 20 for assessments that are unknown to most of Poker Parole’s users. If she chose a different region, she has the option to create a Poker Parole assessment with her own characterisation for indicating that she is sure that her characterisation is the correct one. An assessment created this way is treated the same like assessments created through adding new phrases.

The user interface for **reviewing own assessments** is the same as in Mercato Linguistico and provides the same kinds of information.

### 5.2.2 Customisation of Agora’s Generic Aspects

Poker Parole shares some properties of its aspects with Mercato Linguistico but also differs in some. The similarities and differences in terms of Agora’s core concepts regarding symbolic goods, their characteristics, actions and dividends, and the monetary value of assessments are reflected in the following.

#### 5.2.2.1 *Symbolic Goods*

The symbolic goods under survey are, like in Mercato Linguistico, phrases in Italian language varieties. While in Mercato Linguistico more generally known phrases are gathered, Poker Parole focuses on phrases that most users are not likely to recognise.

#### 5.2.2.2 *Characteristics*

In Poker Parole, the single characteristic that is gathered is the geographical region. Thus, the set of characteristic types is:

$$\{(geographical\ region, number\ with\ relations)\}$$

The characteristic *geographical region* is defined the same way as in Mercato Linguistico.

#### 5.2.2.3 *Actions and their Dividends*

Fewer actions are defined for Poker Parole than for Mercato Linguistico because on the one hand the set of characteristic types has been minimised and on the other hand trading has been deactivated. The following four actions together with their dividends are defined in Poker Parole.

- Unveiling of meaning: The action and its dividend (i. e., none) is defined the same way as for Mercato Linguistico in [Section 5.1.2](#).
- Geographical characterisation: The action itself is the same as for Mercato Linguistico; however, the dividend is different. A dividend is paid to the user if she selects the same geographical region as in the Poker Parole assessment of the phrase. Here, similarity is strict, in contrast to the fuzzy similarity in Mercato Linguistico. In case of a match, the dividend is 20 points for phrases that are difficult to recognise, i. e., at

most 40% of all users recognise its origin, and 10 points for phrases that are not difficult.

- **Geographical assessment:** After a geographical characterisation that did not lead to a dividend payout, the user herself can create an assessment with her own choice of the region. The assessment is then added to the list of the user's own assessments, its monetary value is calculated, and the resulting value is displayed in the user's current amount of play-money. No dividend is paid as part of creating the assessment.
- **Characterisation skip:** When a phrase is displayed to the user, she may skip the characterisation of the phrase and go directly to the phrase of the next round. While this action is dispensable for Mercato Linguistico, it is essential for Poker Parole because the number of users not recognising a phrase have to be counted. Thus, a new custom action is defined for gathering this data explicitly:

**Definition 19** *Given a user  $u$  and a symbolic good  $g$ , the action **characterisation skip** is the tuple  $skip = (u, g)$ .*

No dividend is paid for this action.

#### 5.2.2.4 Monetary Value of Assessments

In Poker Parole, the monetary value is computed as follows: In a first step, the agreement is calculated using a custom similarity function and in a second step, the monetary value is computed.

In contrast to Mercato Linguistico, in which the similarity function values characterisations as neighbouring, descendant, and ascendant geographical regions as similar, the similarity in Poker Parole is strict because the number of exact agreements count. This strictness is important to ensure that users recognise the exact characterisation and not a broader one. The similarity function is therefore defined as:

$$s(a_1, a_2) = \begin{cases} 1 & \text{if } c_1 = c_2 \\ 0 & \text{otherwise} \end{cases}$$

where  $c_i$  is the geographical characteristic in assessment  $a_i$ .

Using this similarity function, the agreement is computed similar to the way it is defined in Agora in [Section 4.2.1](#), with two differences: in Poker Parole, the assessment itself is not included, and characterisation skips are included in the calculation. Thus, the Poker Parole agreement is calculated as follows:

**Definition 20** *Given a state  $(U, G, L, A, D)$  where  $A_a \subset A$  is the set of assessments,  $A_h \subset A$  is the set of characterisations, and  $A_s \subset A$  is the set of characterisation skips, an assessment  $a = (u, g, C, p) \in A_a$ , and a similarity function  $s(h_1, h_2)$ , the **agreement<sub>pp</sub>** :  $A_a \rightarrow [0, 1]$  is defined as:*

$$agreement_{pp}(a) = \frac{\sum_{h_i \in A_{h,g,C,\neg a}} s(h_i, a)}{|A_{h,g,C,\neg a}| + |A_{s,g}|}$$



where

- $A_{h,g,C,\neg a} \subset A_h \setminus \{a\}$  is the set of all characterisations except  $a$  that are compatible with  $a$ ,
- $A_{s,g} \subset A_s$  is the set of all characterisation skips for symbolic good  $g$ , and
- $|\cdot|$  is the cardinality of the set.

This agreement is defined if at least one characterisation or one characterisation skip exists for the symbolic good. Otherwise, it is undefined. The inclusion of characterisation skips as dissimilar characterisations and the exclusion of the assessment of the user herself generally lead to a lower agreement than in Mercato Linguistico. And indeed, a low agreement is what users aim for in Poker Parole. It thus helps users to gain money from their Poker Parole assessments.

The monetary value of a user's assessment is calculated if at least one other user has given or skipped a characterisation for the same symbolic good, otherwise the monetary value is 0. The monetary value follows a different concept than that in Mercato Linguistico. Rather than specifying a maximum value (100) for assessments as in Mercato Linguistico, the monetary value is unrestricted. The more users do not meet the assessment, the more it is worth. Exactly this is needed for gathering phrases that are unknown to many people. Furthermore, a low agreement and a high number of not matching characterisations lead to a high monetary value in Poker Parole. These properties can be reached with the following definition:

**Definition 21** Given a state  $(U, G, L, A, D)$  where  $A_a \subset A$  is the set of assessments,  $A_h \subset A$  is the set of characteristics, and  $A_s \subset A$  is the set of characterisation skips, and an assessment  $a = (u, g, C, p) \in A_a$ , the monetary value of  $a$  is defined as:

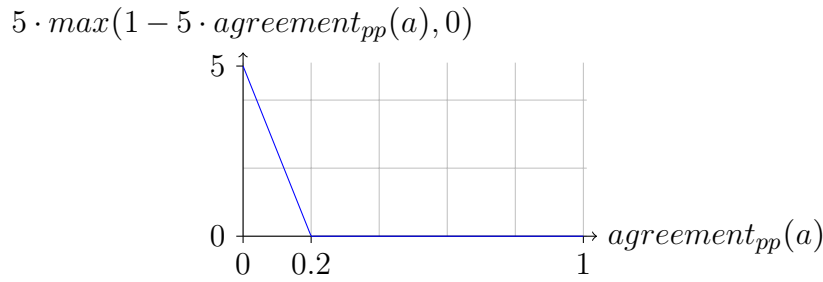
$$value_{pp}(a) = 5 \cdot \max(1 - 5 \cdot agreement_{pp}(a), 0) \cdot (|A_{g,\neg C}| + |A_{s,g}|)$$

where

- $A_{g,\neg C} \subset A_h$  is the set of all assessments on symbolic good  $g$  with characteristics different from  $C$ ,
- $A_{s,g} \subset A_s$  is the set of all characterisation skips for symbolic good  $g$ , and
- $|\cdot|$  is the cardinality of the set.

Figure 11 displays the monetary value without the linear component of the number of not matching assessments. The essential properties of  $value_{pp}$  consist in its decreasing value with increasing agreement and in its low value, e. g., 0, for an agreement of more than circa 0.2.





**Figure 11:** The function  $5 \cdot \max(1 - 5 \cdot \text{agreement}_{pp}(a), 0)$  expressing a part of the definition of the monetary value of an assessment in Poker Parole.

### 5.2.3 Success on Poker Parole: Speculation with Rarely Known Phrases

Given the schema for the monetary value of assessments in Poker Parole, users should aim for phrases that hardly anybody can characterise the same as they themselves do. Therefore, rarely known phrases are gathered in Poker Parole.

Two user strategies could lead to Poker Parole’s exploitation, but both can be detected and remedied. A user might create phrases that do not make sense and therefore cannot be characterised in a sensible way or phrases that make sense but are characterised as being from an irrelevant geographical region. In both scenarios, money is awarded for the assessments although they do not provide value to the platform. Both types of assessments can be detected: The former can be revealed through a) an operator manually reviewing new phrases added to the platform, b) a high number of skipped rounds compared to a low number (probably 0) of characterisations to any region, or c) a button for other users to indicate inappropriate content. Whereas possibility a) requires intervention by the operator and thus is not realistic for massive crowdsourcing, the other possibilities scale with the number of contributors. The latter can be revealed if several other users characterise the assessment as being from the same, “correct” region. Such a detection can be automated as well.

Users performing well in Poker Parole must be specialists for niche dialects or language varieties, opposed to users performing well in Mercato Linguistico who must be generalists for widely known language varieties. Employing these complementary approaches, a wide range of data can be gathered on a unified platform.

## 5.3 USER INCENTIVES

After having introduced the two complementary crowdsourcing applications built with Agora, user incentives for participation are addressed.

Mercato Linguistico and Poker Parole –and crowdsourcing applications in general– can only gather much and high quality data if they provide enough incentives for users to engage in the applications. The incentives which Mercato Linguistico and Poker Parole provide are described in the following.

First, the design as game –or playful application– that entertains and motivates the user is an incentive in itself. This also includes the gaming aspects (play-)money, rankings of the best users, and game rounds as further incentives. To avoid the user’s discouragement, she can skip phrases or characterizations she does not know or want to give.

Research suggests that user incentives can be most effective when incorporating the positive *social facilitation* effect and avoiding the negative *social loafing* [135]. Social facilitation describes that users tend to solve simple tasks better with someone else watching them than without supervision. Social loafing describes that users make less effort to solve tasks when working in a group than alone. Therefore, the following aspects are important and have been incorporated into Mercato Linguistico and Poker Parole:

- The accomplishments of individual users should be shown prominently to avoid social loafing — ranking lists show the top performing users.
- Other users should be able to evaluate each user’s contribution in order to support social facilitation — all entered characteristics are displayed in the results view for a phrase.
- The unique value of each user’s contribution should be highlighted to avoid social loafing — a summary is shown to the user after the last round highlighting her own actions during the played rounds.

By incorporating such social psychological incentives, users tend to contribute more data and return to the Web platform [36].

Also, performing well on a market is an incentive in itself. This is true for financial markets like Wall Street and also for markets embedded in board games, like monopoly [47]. For social media applications built with the market-based operating system Agora, it is not yet proven but highly likely to be true as well. Especially for power-motivated and achievement-motivated stereotypes (see Section 4.3.1 and [78]), their interest in performing better than the crowd is apparent in all kinds of markets. Furthermore, each kind of market involves a gaming dimension in itself as traders are playing with each other with their speculations in order to get the best performance on the market. These incentives also are apparent in the market-like applications Mercato Linguistico and Poker Parole, in which the user’s play-money depends on her speculation as well as on the other users’ assessments.

Concerning language, in all cultures there is a considerable interest in language issues and in reflecting on one’s own language variations. People interested in their own language are likely to participate in Mercato Linguistico and Poker Parole just for seeing what others disclose on the platform, both phrases or sentences they do not know and assessments they are not aware of. Also if a user were not attracted by games in general but interested in language variations, she might still consider participating for the sake of her interest in language.

## 5.4 THE COLD START PROBLEM

The “cold start” problem describes the situation at the beginning of gathering user data in which not enough data or participants exist for an application to work in the way it is designed. The effects of the cold start problem are two-fold, the lack of data and the lack of participants.

A cold start problem exists for the phrases themselves: at least some phrases have to be present in the application so that participants can engage. Seed data must therefore be gathered and entered into the platforms before launching them to the public. For Mercato Linguistico and Poker Parole, around 800 phrases in Italian language varieties have been gathered together with the geographical region and included before publishing the applications on the platform metropolitalia. Although gathering seed data can be time-consuming, it is a precondition for the operation of the applications.

If only few users participate so far, for example, in a GWAP, other participants can be annoyed by the low amount of possible interaction or even cannot play at all, in the case of simultaneous multi-player games like the ESP game [184] (for details on the ESP game see [Section 2.1.4](#)). Common approaches to remedy this situation consist in recording games with players and replaying them to other players or in creating bot players having some kind of artificial intelligence [184, 185, 205]. Mercato Linguistico and Poker Parole are no simultaneous multi-player applications and therefore the effect of few participants playing simultaneously does not hinder participation and no bots or replays have to be introduced.

However, if no feedback can be provided to participants because of a lack of characterisations, they may be annoyed. This is an effect of the lack of data. To counteract such annoyance, phrases with enough characterisations alternate with phrases with too few characterisations that need more characterisations.

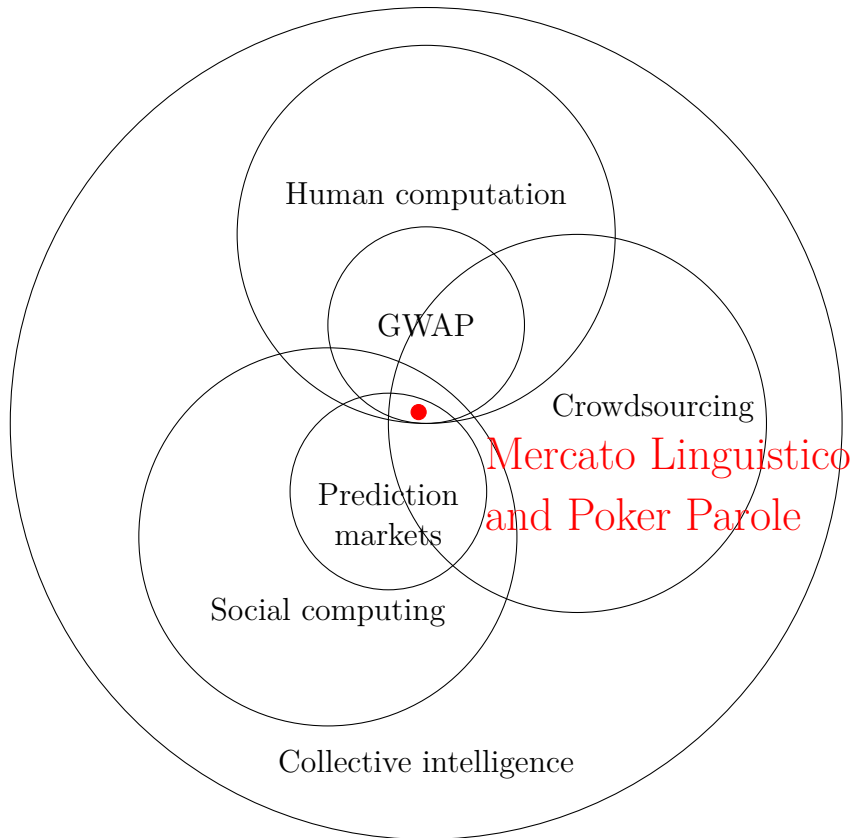
The cold start problem for lacking data is also investigated in the context of recommender systems, in which the problem consists in providing recommendations for new users. Methods for predicting recommendations using machine learning algorithms [177] or using training data provided by users [138] have been investigated, but cannot reliably provide profound recommendations in all cases. For metropolitalia, a machine learning approach is conceivable for predicting, e.g., the geographical region where a phrase is spoken. The algorithm could be based on characterisations of other, similar phrases which contain one or more common words that are specific for the region. However, in order to successfully build a machine learning algorithm for Italian language varieties, a decent amount of data has to be gathered in the first place, which is among the goals of metropolitalia. Therefore, it does not solve the seed data problem.

Thus, the manual gathering of phrases and their geographical characteristics was the only applicable solution for the cold start problem. By that, a dataset was available on which basis feedback for geographical characterisations and assessments could be given to participants. For other characterisations and

word selection, no feedback is given for the first participant. Instead, a hint is shown that the entered data is scored when more data is available.

## 5.5 CLASSIFICATION IN THE FIELD OF COLLECTIVE INTELLIGENCE

Mercato Linguistico and Poker Parole can be classified in the research fields that have been introduced in [Section 2.1](#). The classification is depicted visually in [Figure 12](#).



**Figure 12:** Mercato Linguistico and Poker Parole depicted in the diagram of collective intelligence and related fields, originally introduced in [Figure 1](#) in [Section 2.1](#).

Both social media applications are clearly crowdsourcing applications as they gather data from many people using an open call for participation. Also, social computing techniques are applied in both as users see other users' characterisations and the collective results. Furthermore, Mercato Linguistico and Poker Parole are applications of human computation because an algorithm is conceivable that characterises dialect phrases geographically and according to the speaker's social characteristics. Such an algorithm could be trained by employing data gathered in the two applications.

Furthermore, Mercato Linguistico and Poker Parole can be considered being GWAPs in addition to human computation applications. Salen and Zimmerman [175] investigate different definitions of games and conclude that “a game is a system in which players engage in an artificial conflict, defined by rules in a quantifiable outcome” [175]. Several components are apparent in this definition. All of them can be seen as met by Mercato Linguistico and Poker Parole: They are systems in which one or more players actively engage. The so-called conflict is the achievement of assessments with a high monetary value. The conflict is artificial, opposed to occurring in the real world. The player is restricted by rules. And the outcome is quantifiable through the play-money a player obtains. Therefore, Mercato Linguistico and Poker Parole are games and –as they belong to the area of human computation– GWAPs.

Besides the conceptual game model that accounts for the applications being a GWAP, also other game design patterns like game rounds or play-money are incorporated into the two social media applications to lead to a more “gamified” experience. Therefore, gamification techniques (see [Section 2.3.2](#)) can be seen as applied as well.

Also features of prediction markets are incorporated. In that users estimate the agreement proportions of assessments, they predict how the underlying phrase is characterised by all users in a market-based way and thus create a prediction with a numerical value. This is similar to prediction markets in which the prediction happens through the purchase and sale of stock options. Different from prediction markets, the predicted event cannot be verified easily at some time in the future and the market component encompasses the aggregation of results in addition to trading in terms of purchases and sales.

## 5.6 POTENTIAL OF AGORA IN ART HISTORY

The complementarity of Mercato Linguistico and Poker Parole can be exploited in completely different areas for field research. As an example, its possible application in art history is sketched. This area has been chosen because the art history project ARTigo [25, 112, 214, 233] is also developed within the Play4Science project [113, 264] and therefore knowledge about crowdsourcing applied to art history can be exploited.

Two similar complementary social media applications like Mercato Linguistico and Poker Parole are conceivable and would yield new insights into the perception of artworks. The symbolic goods traded with would be artworks and the characteristics assessed could be the artist, style, and epoch. Other than changing the graphical user interface appropriately for displaying images of artworks instead of phrases and choosing the characteristics appropriately, the software for running such artwork-oriented applications would stay the same. Agora can be used as it is.

In an artwork market-based application similar to Mercato Linguistico, users could upload images of artworks as symbolic goods and users would speculate on the characteristics of artworks and perform well if other users assign the

same characteristics. For example an image displaying the painting “Mona Lisa” would be well recognised to be created by the famous painter Leonardo da Vinci. This market would unveil what art generalists recognise in artworks. Users would learn about the artists, styles, and epochs of artworks by speculating on the market.

In a complementary artwork game similar to Poker Parole, users would speculate that others do not know the characteristic of an artwork. Many artworks can be mistaken for something else than what they are. For example, the authors of impressionist paintings can be confused. Such confusions are sometimes typical of artworks or painters. An artwork game in the style of Poker Parole would unveil valuable information on both the artworks and the players, which could be used for realising or improving an artwork search engine.

Not only people wanting to learn about art are addressed, also experts seeking information about the recognition of artworks are catered for. Therefore, both who are the experts for specific areas of art history and what only these experts recognise would be unveiled in such applications. Participating in these applications can be fun because the characteristics revealed are interesting for users as they are expert knowledge.

The complementarity of media in the style of Mercato Linguistico and Poker Parole is likely to be exploitable in other areas, as for all areas both general and expert knowledge exist.

## 6

EVALUATION OF DATA  
GATHERED WITH MERCATO  
LINGUISTICO AND POKER  
PAROLE

Aspects of the content of this chapter have been published in [24, 109, 111]: the introduction of the chapter is partly based on [24], Section 6.2 and Section 6.3 extend parts of [24, 111], and Section 6.4 is partly based on [24, 109, 111].

The platform *metropolitalia* has been publicly available since August 2012. As its first application, *Mercato Linguistico* was deployed at that time for getting first insights into its acceptance. In April 2013, *Poker Parole* was published for complementing *Mercato Linguistico* with gathering rarely known phrases that are not gathered with *Poker Parole*.

This chapter presents an evaluation of data gathered on *metropolitalia* within the two applications. The evaluation is based on the data gathered until the end of October 2013. It examines data gathered within the relatively short time frame of 15 months, compared to the envisaged operation of the platform of many years. The amount of data gathered in this time frame is relatively small because only a limited amount of participants could be attracted. This is due to the small amount of advertisement and due to the Italian-speaking project members not taking care enough of making *metropolitalia* popular. The evaluation given here focuses on the data gathering mechanisms based on *Agora*. Examining the gathered data from a linguistic perspective remains for experts of Italian language varieties in the area of Romance linguistics. Due to the small amount of gathered data so far, an evaluation of trading cannot be covered as part of this thesis and remains for future work.

For gathering data, *metropolitalia* faces the same problem as every platform that is first published on the Web: It is unknown and it has to attract users. The *Play4Science* project did not have financial means for advertising or for offering financial rewards of platform participants. Therefore, especially social media channels were employed for making the platform known. Personal contacts of the members of the project were informed, blog articles were published at Italian Web blogs, a blog accompanying the platform was established, a Facebook site was set up and promoted, a Twitter account was established, and the project was promoted to native Italian researchers on scientific conferences. Interest in the own language, the design as playful applications, feedback to users, and the market mechanisms act as incentives (as described in Section 5.3). No real money or prices are awarded to participants, e.g., for having the first position in the list of the best users.

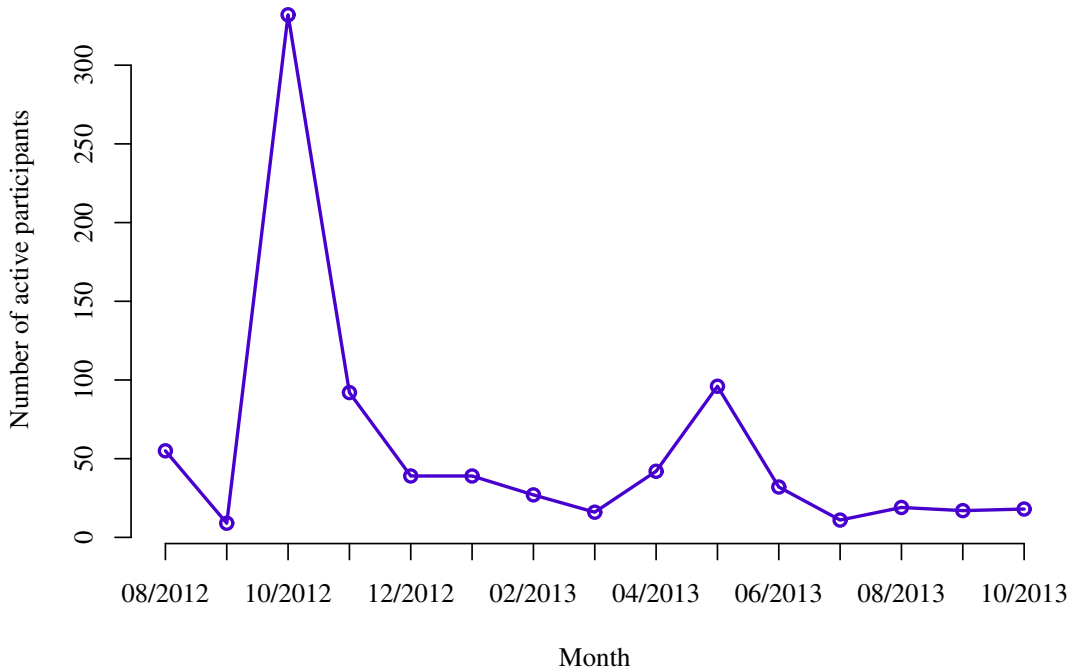
Test data from members of the Play4Science group has been excluded for this evaluation in order to not influence the results.

This evaluation is divided into the following parts:

- the temporal development of participants and gathered data,
- a quantitative breakdown of data gathered in total, per user, and per phrase,
- an analysis of the assessment's estimated agreement proportion, and
- data quality and usefulness for research on Italian language varieties.

## 6.1 TEMPORAL DEVELOPMENT

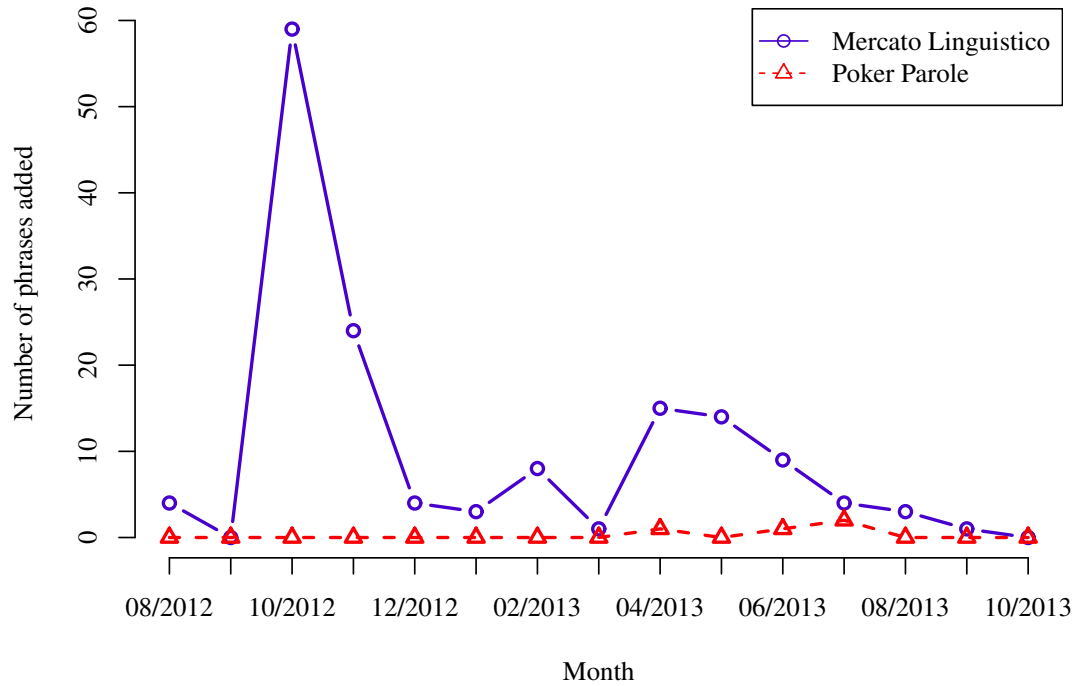
Data has been gathered continuously over the whole 15 months period with the two applications Mercato Linguistico and Poker Parole. An analysis of the temporal development can yield insights into the acceptance of the applications and their individual data gathering methods. In the following, the most important temporal developments of numbers are investigated: active participants, added phrases, gathered characterisations, and gathered assessments.



**Figure 13:** Active participants on metropolitalia per month.

Figure 13 shows the temporal development of the number of active participants on metropolitalia. Active participants are users who submitted at least one characterisation or created a phrase within a time frame. Participants of Mercato Linguistico and Poker Parole are included here. The first big peak during October 2012 stems from an article by a member of the Play4Science





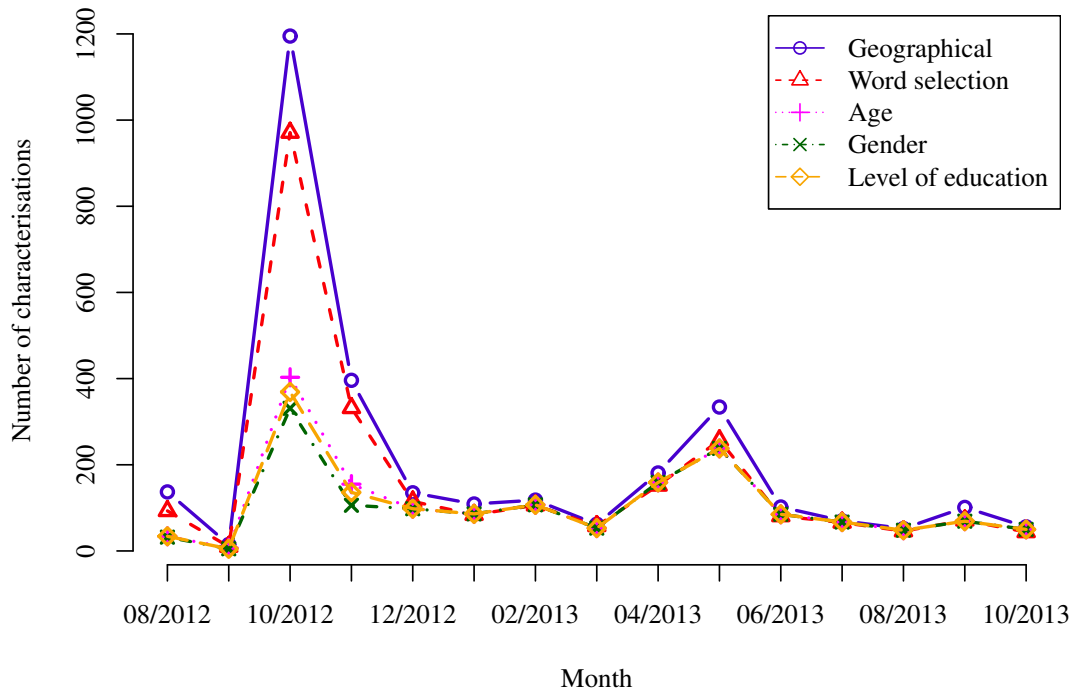
**Figure 14:** Number of new phrases added to Mercato Linguistico and Poker Parole by users per month.

project published on an Italian blog with many readers [252] and the social media resonance related to this article on Facebook and Twitter. Also in the following months, this article attracts participants. The second peak in May 2013 is from a news item created on metropolitalia’s Facebook page [261] which was shared among many friends.

As it can be seen, social media could attract many visitors to metropolitalia. However, the enduring engagement of participants and a wider publicity apart from social media could not be reached so far. This may be possible through a sequential publication of articles in order to have a continuous flow of users coming to the platform.

In Figure 14, the number of new phrases that have been added to Mercato Linguistico and Poker Parole by users are displayed. These curves follow a similar trend as the curve of active participants in Figure 13. The difference between the trends of the curves is mainly apparent in February and April 2013, which probably is a normal fluctuation.

Figure 15 displays the number of gathered characterisations over time. The curves of the individual characterisations follow the same trend as the trend of active participants on metropolitalia (see Figure 13). The number of geographical characterisations is always the highest of all characterisations because a geographical characterisation must be given by users before being able to give other characterisations. From August 2012 to November 2012, the second highest number is the number of characterisations for selecting relevant words. In these months, the graphical user interface for gathering the characterisations age, gender, and level of education appeared to be confusing for users leading to less such characterisations than those for selecting relevant words. The user

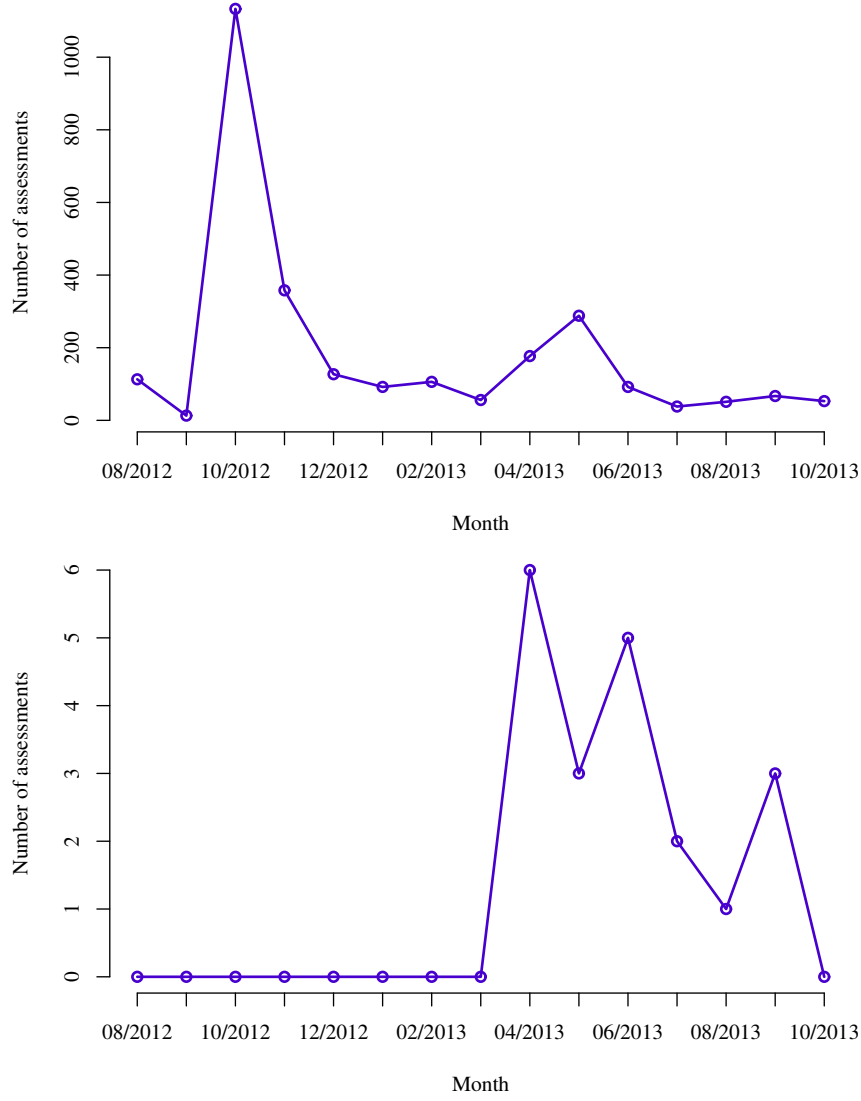


**Figure 15:** Number of characterisations gathered with Mercato Linguistico per month.

interface was changed from a graphical slider to usual radio buttons, which are common on many other websites. Since December 2012, the amounts of gathered data for characterisations except for geographical characterisations are almost the same for each month.

In [Figure 16](#), the temporal development of the number of assessments gathered with Mercato Linguistico and Poker Parole is displayed. The much lower number of assessments gathered with Poker Parole, compared to the number of assessments gathered with Mercato Linguistico, stems from two main reasons: First, in the navigation menu on the left part of the website, Poker Parole is listed below Mercato Linguistico, leading to more users trying the application listed first, i.e., Mercato Linguistico. Second, the main web page of metropolitalia prominently displays a map of Italy with phrases displayed as speech bubbles on the map. Upon clicking on this map, the Mercato Linguistico application is started. Therefore, more first time users participate on Mercato Linguistico than on Poker Parole and more assessments are gathered with the former than the latter application.

The analysis of the temporal development of data gathered with the two applications Mercato Linguistico and Poker Parole shows that users participate and engage on the platform mainly when they read about it, especially on social media like weblogs or Facebook. Without current direct publications to a mass of users, only few users can be led to the platform. This is especially true for young and unknown websites.



**Figure 16:** Number of assessments gathered with Mercato Linguistico (top) and Poker Parole (bottom) per month.

## 6.2 QUANTITATIVE BREAKDOWN

Besides the temporal development of the data gathering process on metropolitalia, the total amount of gathered data is of interest. The following sections provide a quantitative breakdown. First, the total amount of gathered data is described, second, the gathered data per user, and finally, the gathered characterisations per phrase.

### 6.2.1 Total Amount of Gathered Data

The amount of gathered data within the time frame of 15 months can be seen in [Table 2](#). In total, 857 users visited the platform for playing 4940 rounds of Mercato Linguistico and 160 rounds of Poker Parole. Within these, 3206 times a geographical location was assigned by the user, 2812 assessments (that

	Mercato Linguistico	Poker Parole
Participants	840	40
Played rounds	4892	146
Geographical characterisations	3063	112
Assessments	2764	20
Word selections	2487	-
Age characterisations	1338	-
Gender characterisations	1218	-
Level of education charact.	1261	-
Created phrases	149	4
Participants who created phrases	82	4

**Table 2:** Total amount of gathered data. Characterisations with value “unknown” were not counted.

is, geographical characterisations with estimated agreement proportions) have been created, 2508 times one or several words were selected as being relevant for the geographical characterisation, 1342 / 1222 / 1261 social characterisations have been produced for the social attributes age / gender / level of education, and 154 phrases have been created.

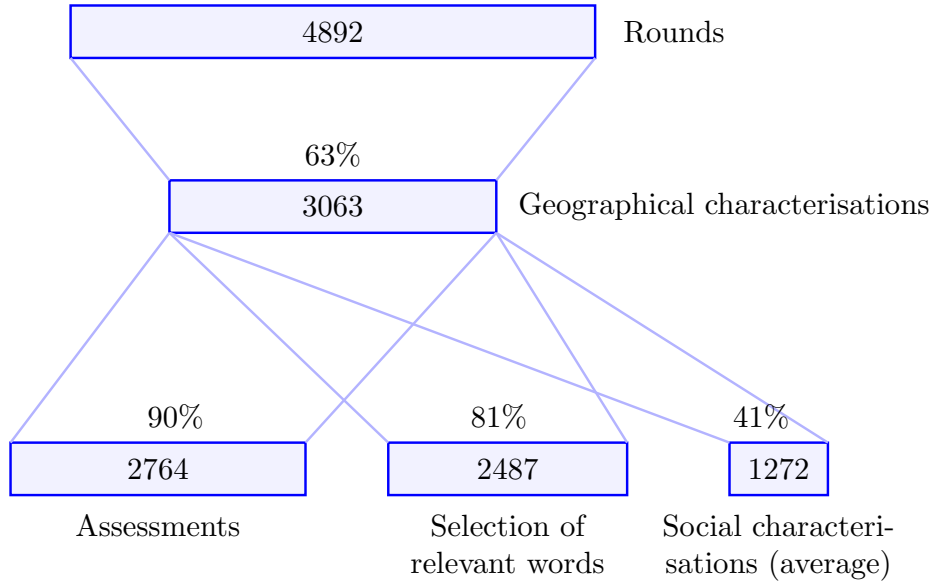
The amount of data gathered with Poker Parole is still quite low, due to its recent publication and reasons outlined in [Section 6.1](#). Therefore, only data gathered with Mercato Linguistico is evaluated in the following. An evaluation of Poker Parole’s data remains for future work.

During one round of Mercato Linguistico, not every action is performed by users, as depicted in [Figure 17](#). In 63% of all rounds a geographical region is chosen, in 38% of all rounds the phrase is skipped. The reason for the relatively high percentage of skipped rounds probably is that the user did not know the phrase well enough to estimate a geographical region of its occurrence. This is natural and was foreseen, giving users the option to skip rounds.

In 90% of the rounds in which a geographical characterisation is undertaken, an assessment is created. This high number indicates that users are confident in giving estimations on the agreement proportion, a finding that encourages to employ Agora’s market-based approach of assessments in further applications.

The high percentage of word selections, 81%, are probably due to the relative ease of the task. On the one hand, native Italians should recognise the words that are not standard Italian, on the other hand, users can display a translation of a phrase to standard Italian, with which it is often clear which word(s) of the phrase are specific for this language variety.

The lower percentage of social characterisations (51% on average) bear evidence that some phrases cannot be characterised in the social attributes age, gender, and level of education from a linguistic point of view. Furthermore, completing these steps is optional.



**Figure 17:** Number of actions a user performs in Mercato Linguistico. The number of social characterisations is the average of the number of age, gender, and level of education characterisations.

Not only the data gathering process on the characterisation side can be seen to be successful, also the possibility for users to add new phrases or sentences led to 154 new phrases that were contributed to metropolitalia by users.

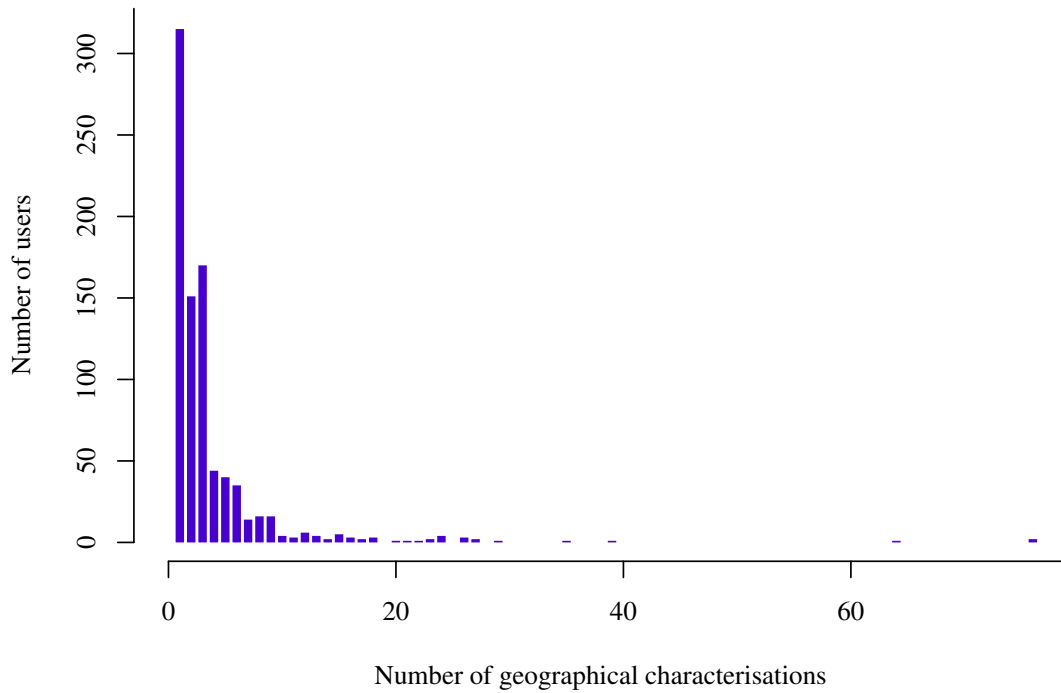
To summarise, many characterisations and new phrases have been gathered during the first 15 months of Mercato Linguistico. This represents a good start of the platform on which foundation more data is to be gathered.

### 6.2.2 Gathered Data per User

Besides the temporal development and the total amount of gathered data, it is interesting to investigate how many data each individual user contributed.

Regarding the number of phrases created by participant, 10% of all users who played at least one round of Mercato Linguistico or Poker Parole contributed new phrases (82 out of 840 respectively 4 out of 40, see Table 2). This indicates, but not very strongly, that the incentives for adding phrases are good enough. It is a parallel conclusion drawn from the comparison with other social media sites: Compared to the number of all users visiting the platform metropolitalia (3749 according to the web analytics tool Piwik [263]), the percentage of phrase contributing users is 0.023%. This is on par with the contribution percentage of users on other social media sites. For example, Wikipedia estimates that 0.02-0.03% of all visitors actively contribute to Wikipedia [274].

In Figure 18, the distribution of the number of geographical characterisations generated by single users on Mercato Linguistico is displayed. Many users (316) created only one geographical characterisation and then left metropolitalia. Still a lot of users (322) created two or three geographical characterisations, which is usually the case when playing for three rounds until the feedback page



**Figure 18:** Distribution of the number of geographical characterisations generated by single users on Mercato Linguistico.

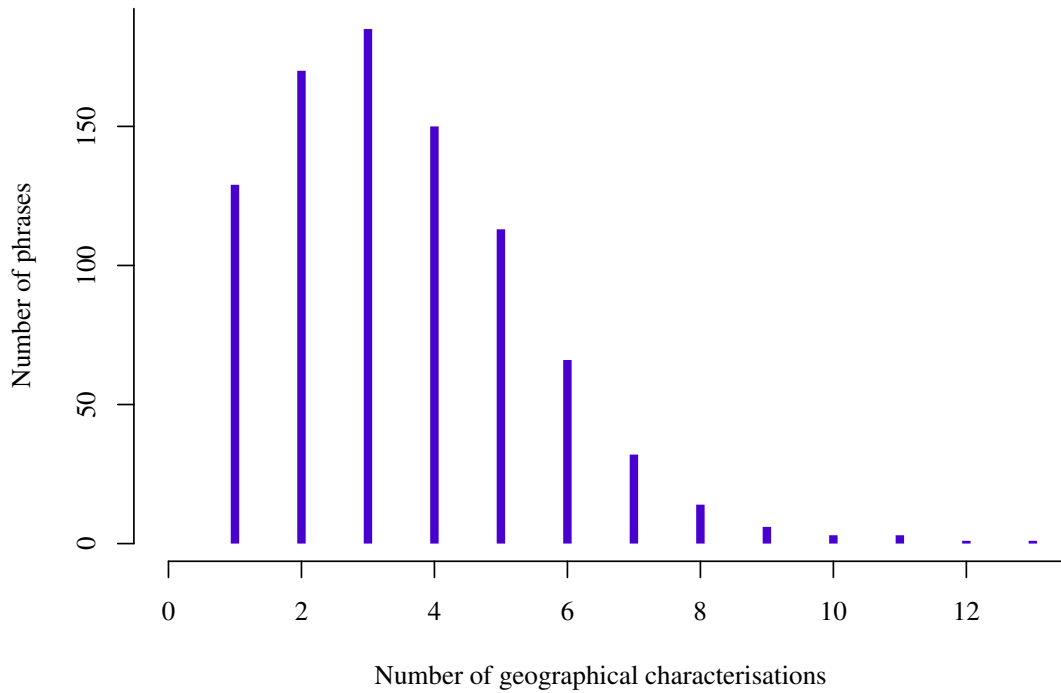
is displayed. Note that users may skip phrases they do not know and thus users may reach the feedback page with only one or two geographical characterisations. Few users participated for a long time on Mercato Linguistico, creating more than 60 geographical characterisations. Besides these few active users, many short term users contribute much and valuable data.

The distributions of the number of assessments and of other characterisations generated by single users are nearly the same as the distribution for geographical characterisations in Figure 18. These distributions are included in Appendix A.1.1 for reference.

### 6.2.3 Gathered Characterisations per Phrase

For an evaluation of individual phrases, the number of characterisations per phrase is important to know. Figure 19 displays the distribution of geographical characterisations per phrase on Mercato Linguistico. The distributions for the other characterisations can be found in Appendix A.1.2. Most phrases have around three characterisations. This low number of data makes it difficult to analyse the phrases from a linguistic point of view. For some of the phrases, up to 14 characterisations have been gathered. These phrases are mostly gold standard phrases that are included for verification purposes (see Section 4.5.2).

What number of characterisations is needed for statistically significant data depends on the type of characterisation. The more choices users have, the more data is needed usually. For example, geographical characterisations usually need more data than social characterisations like age, gender, or level

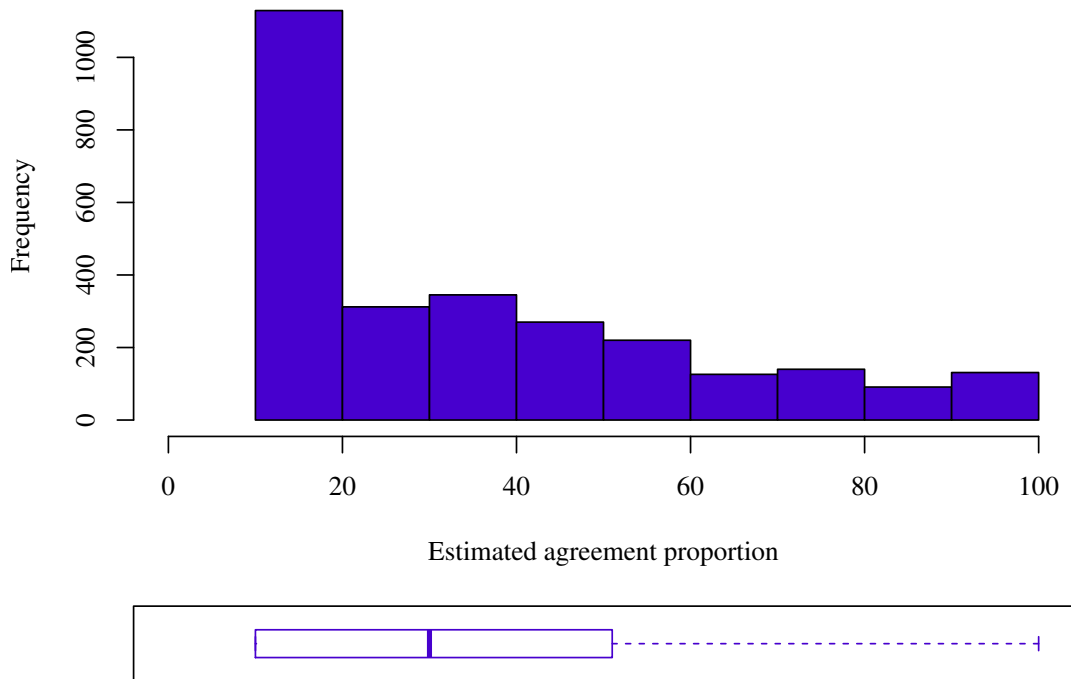


**Figure 19:** Distribution of the number of geographical characterisations per phrase on Mercato Linguistico.

of education because of their richness in choice. Therefore, social characterisations may already yield results that are supported by statistical measures while geographical characterisations only yield vague hints.

### 6.3 EVALUATION OF THE ASSESSMENT'S ESTIMATED AGREEMENT PROPORTION

As a central concept of Agora, the assessment and its estimated agreement proportion need to be evaluated. The distribution of the estimated agreement proportion of assessments given by the users in Mercato Linguistico is displayed in [Figure 20](#). The whole allowed range between 10 and 100 is present. Interestingly, most users estimate the agreement proportion relatively low, the median is 30 and 50% of the proportions are less than ca. 50. Reasons could be that users consider their characterisation being relatively unknown to other users or that users may not be very confident in their choice. Another reason could be that the default value given in the graphical user interface is 10 and thus many users leave it at this default value. This possibility could be excluded by a correction of the user interface in the next version, for example by adjusting the default value. A reasonable default value could be derived by taking the mean of the estimated agreement proportions of assessments with a high monetary value, i. e., a low difference between the estimated agreement proportion and the calculated agreement. For example, the mean of assessments with a



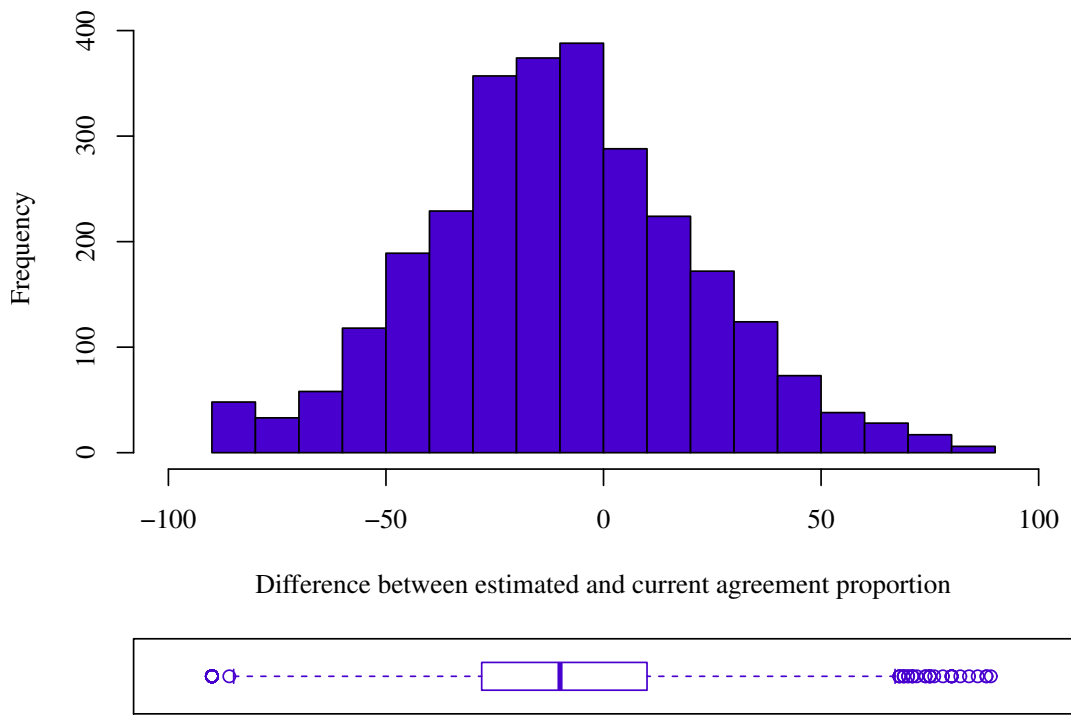
**Figure 20:** Histogram with boxplot of the estimated agreement proportion of assessments as specified by users in Mercato Linguistico.

difference of at most 10 is 36 (see Appendix A.1.3), which may be a reasonable default value.

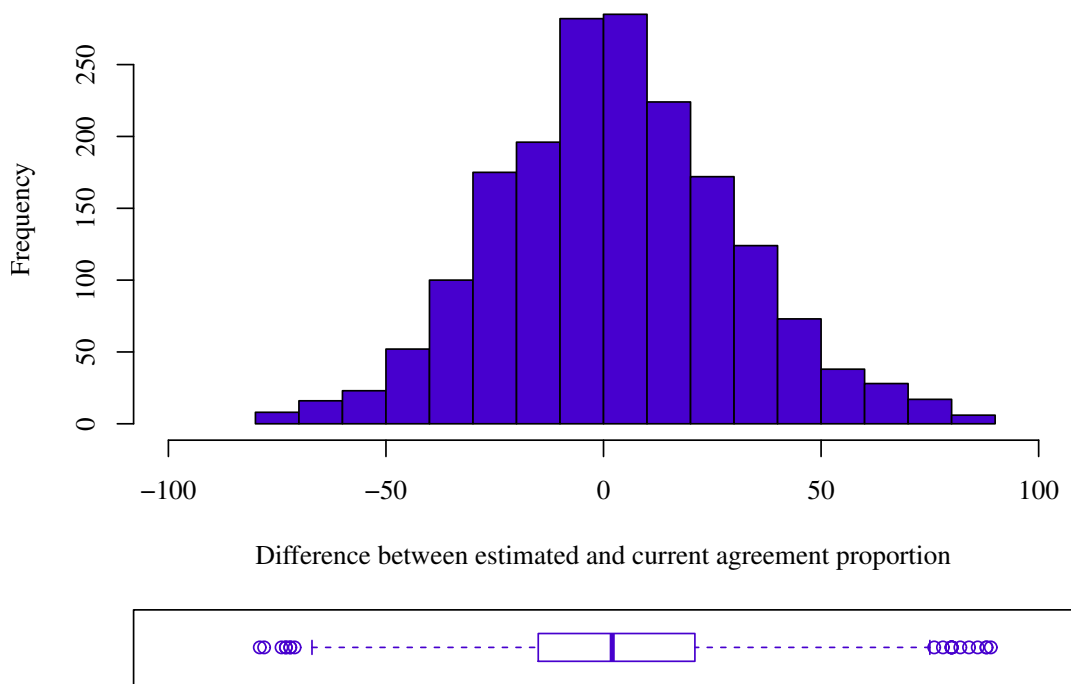
The estimated agreement proportion can now be compared to the calculated agreement as of the current data status. Figure 21 shows the distribution of the differences between the first and the latter value. If the estimated agreement proportion is greater than the calculated agreement, the value in the distribution is  $> 0$ . If the former is smaller than the latter, the value in the distribution is  $< 0$ . Values  $> 0$  can be considered overestimations and values  $< 0$  can be considered underestimations. The distribution shows that in general the agreement proportion is very slightly underestimated by users. However, this may be due to the high number of assessments with an estimated agreement proportion of 10, which may have been influenced by the user interface. And indeed, if all such assessments are excluded, the distribution yields a small overestimation, as Figure 22 depicts. Then, the distribution is balanced.

To sum up, the following conclusions can be derived for assessments from the evaluation. First, assessments are widely accepted by users, as 90% of all geographical characterisations are extended to assessments with an estimated agreement proportion. This is promising because the user acceptance is critical to deploy market-based assessments in further applications. Second, users are relatively good at estimating the agreement proportion. For 53% of the assessments the difference between the estimated and the calculated agreement is less than 20 percentage points (for Figure 22). It remains to evaluate the quality of assessments for linguistic field research, which is done in the following.





**Figure 21:** Histogram of the difference between the estimated agreement proportion and the calculated agreement of assessments in Mercato Linguistico. Values  $> 0$  are overestimations and values  $< 0$  are underestimations.



**Figure 22:** Histogram of the difference between the estimated agreement proportion and the calculated agreement of assessments in Mercato Linguistico for which the estimated agreement proportion is not 10. Values  $> 0$  are overestimations and values  $< 0$  are underestimations.

Geographical region	Estim. a.p.	Calc. a.	Monet. value
North	27%	50%	12
North	81%	50%	2
North	100%	50%	0
North>Laives	34%	45%	56
North>Lombardia	44%	50%	86
North>Rimini	10%	45%	1
South>Calabria	45%	13%	2
South>Matera	51%	24%	5
South>Taranto	35%	18%	31

**Table 3:** Gathered geographical assessments for the phrase “Se non la smetti ti do una sberla.” (in English “If you don’t stop it, I’ll box your ears.”) including the estimated and the calculated agreement and the monetary value of the assessment.

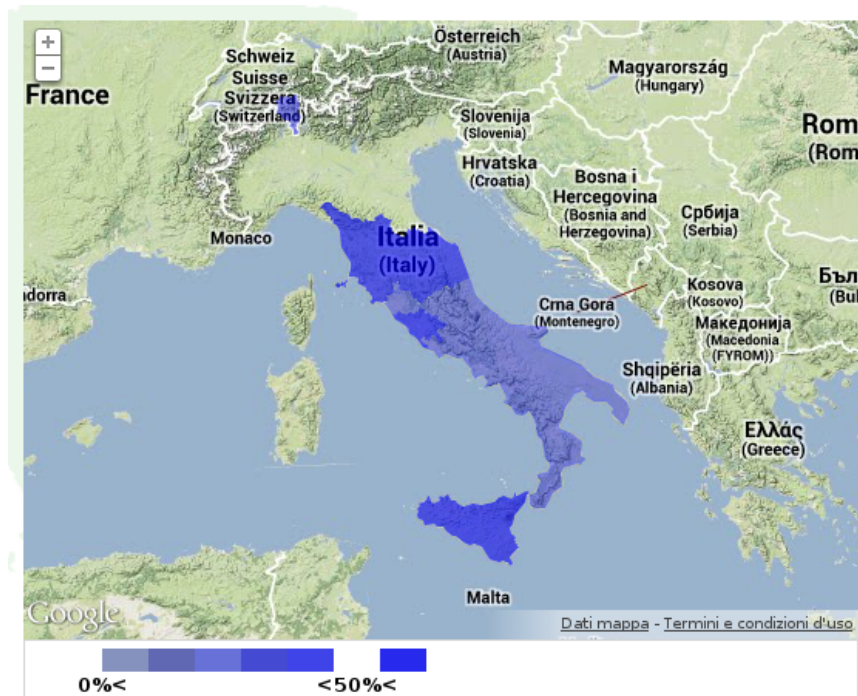
## 6.4 QUALITY OF GATHERED DATA FOR LINGUISTIC FIELD RESEARCH

For linguistic field research, the quality of gathered data is important besides its quantity. In this thesis, two examples are discussed in depth. A more thorough evaluation remains for experts of Italian variety linguistics.

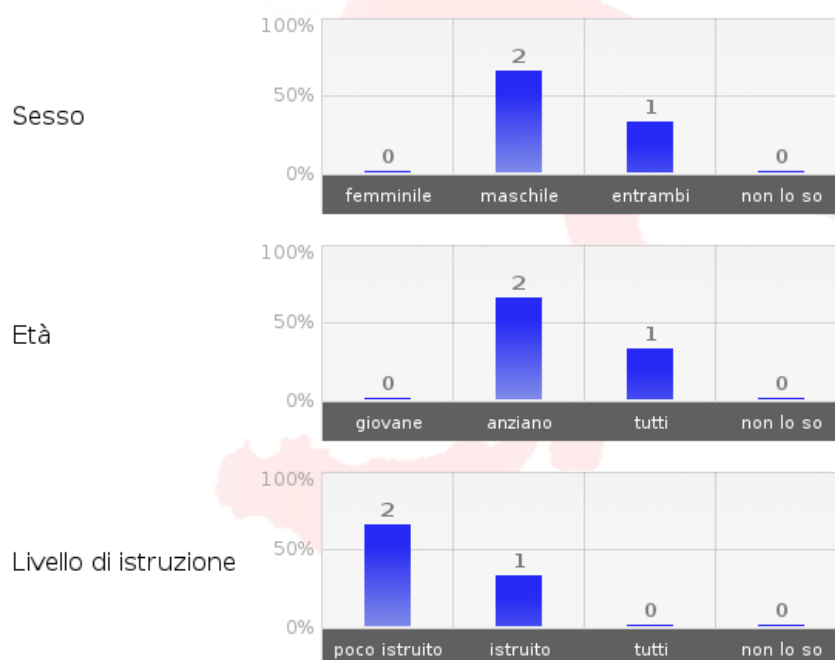
In [Figure 23](#), the results for a phrase as displayed on metropolitalia are shown. The phrase “Mio figlio è proprio un femminaro!” (in English: “My son really is a womaniser!”) is characterised to be spoken more in the south of Italy (see the coloured map), the speaker is characterised as male, older, and less educated (see the three bar charts), and the selected relevant word is “femminaro”, a vernacular word for a womaniser. Though only six users characterised the phrase so far, a tendency to the use in the center and south of Italy can be seen. And according to a native Italian speaker knowing this word, it is well known in Sicily (island in the south of Italy).

It is also of interest how precise the estimated agreement proportions are. It can be seen that they are of varying accuracy. In [Table 3](#), the geographical assessments for the phrase “Se non la smetti ti do una sberla.” (in English “If you don’t stop it, I’ll box your ears.”) are shown. Some assessments meet the calculated agreement quite well (as the ones for the northern Italian regions Lombardia and Laives) and are worth a lot, while others do not estimate the agreement of users on the platform well (as the ones for the southern Italian regions Calabria and Matera).

Not only the match to the community’s opinion on the platform is of interest, also the match to the findings of researchers. Here, according to researchers of the Italian language, the word “sberla” originally spread from northern Italy [\[17\]](#). The mean of the estimated agreement proportions for the geographical region “North” (without subregions) is 69%. The mean suggests that the phrase



### L'opinione dei giocatori:



### Parole rilevanti:

Mio figlio è proprio un femminaro !  
 0 0 0 0 0 5 0 di 5

2 su 6 (33%) puntano su: Centro  
 1 su 6 (16%) punta su: Sud  
 1 su 6 (16%) punta su: Sud>Sicilia  
 1 su 6 (16%) punta su: Sud>Lazio>Roma>Roma  
 1 su 6 (16%) punta su: Ticino>Prosito

**Figure 23:** metropolitalia platform displaying the data gathered for the sentence “Mio figlio è proprio un femminaro!” (in English: “My son really is a womaniser!”)

tends to be used more in the northern part of Italy. Therefore, the estimated agreement proportions support the scientific distribution.

Furthermore, this mean of the estimated agreement proportions is higher than the calculated agreement of 50%. As the phrase originally spread from northern Italy, the mean of the estimated agreement proportions yields a closer estimation than the pure characterisations. If only the broad geographical regions “North” and “South” were taken into account without subregions, the broad distribution would be that 6 out of 9 assessments, i. e., 66%, were “North” and 3 out of 9 assessments, i. e., 33%, were “South”. This value is close to the mean of the estimated agreement proportions. However, the users who chose a smaller region did this deliberately, and this should be reflected in the distribution. An adjustment of the similarity function and its constants may yield a more precise value for the calculated agreement.

The estimated agreement proportion also gives hints about the perception of users. The users who proposed the highest estimated agreement proportions are two of the three users who chose “North”. These users can be seen as being quite sure that almost everyone else on the platform would also characterise the phrase as being from “North”. This shows that they perceived that it is well known in whole Italy that the phrase is common in northern Italy. Interestingly, most of the other users gave smaller estimated agreement proportions around 40%. So they perceived the chosen origin of the phrase to be less known amongst Italians. Such data may lead to new findings in the area of language perception. It is interesting to see how these influence research in the area of Italian language varieties, which remains to be evaluated by Romance linguists.

As it can be seen on these examples, the quality of the gathered data is convincing, and the assessments’ estimated agreement proportions can provide more precise distributions than “usual” characterisations. Till now, the relatively low amount of data hinders a more elaborate data evaluation. When metropolitalia will receive more attention in the Italian community, it will also gather a higher quantity of data especially resulting in more assessments per phrase being available. An expert in the area of variety linguistics on the Italian language starts seeing metropolitalia as a promising step towards new methods of gathering data, emphasising its significance [114].

## Part III

# E-LEARNING



Aspects of the content of this chapter have been published in [108]: especially Section 7.1, Section 7.1.1, Section 7.1.2, Section 7.2.1, and Section 7.4 are derived from [108].

After having shown in the previous part of the thesis how crowdsourcing techniques can be used for linguistic field research, this part is concerned with the combination of crowdsourcing and e-learning.

To this aim, a game-like e-learning tool, called *Termina* (derived from Latin “terminus”, meaning “technical term”), has been conceived as part of this thesis. In *Termina*, lecturers define a set of concepts for a certain topic, usually the topic of a university course. Then, students propose associated terms for the given concepts and lecturers classify associated terms into *close* and *far* for providing valuable feedback to students. Collectively, both contribute to the construction of association maps. This thesis proposes the term “association map” for a simplified version of concept maps [151] with only one proposition type, namely the association type. Thus, association maps are visual representations of concepts with unlabelled relationships between concepts.

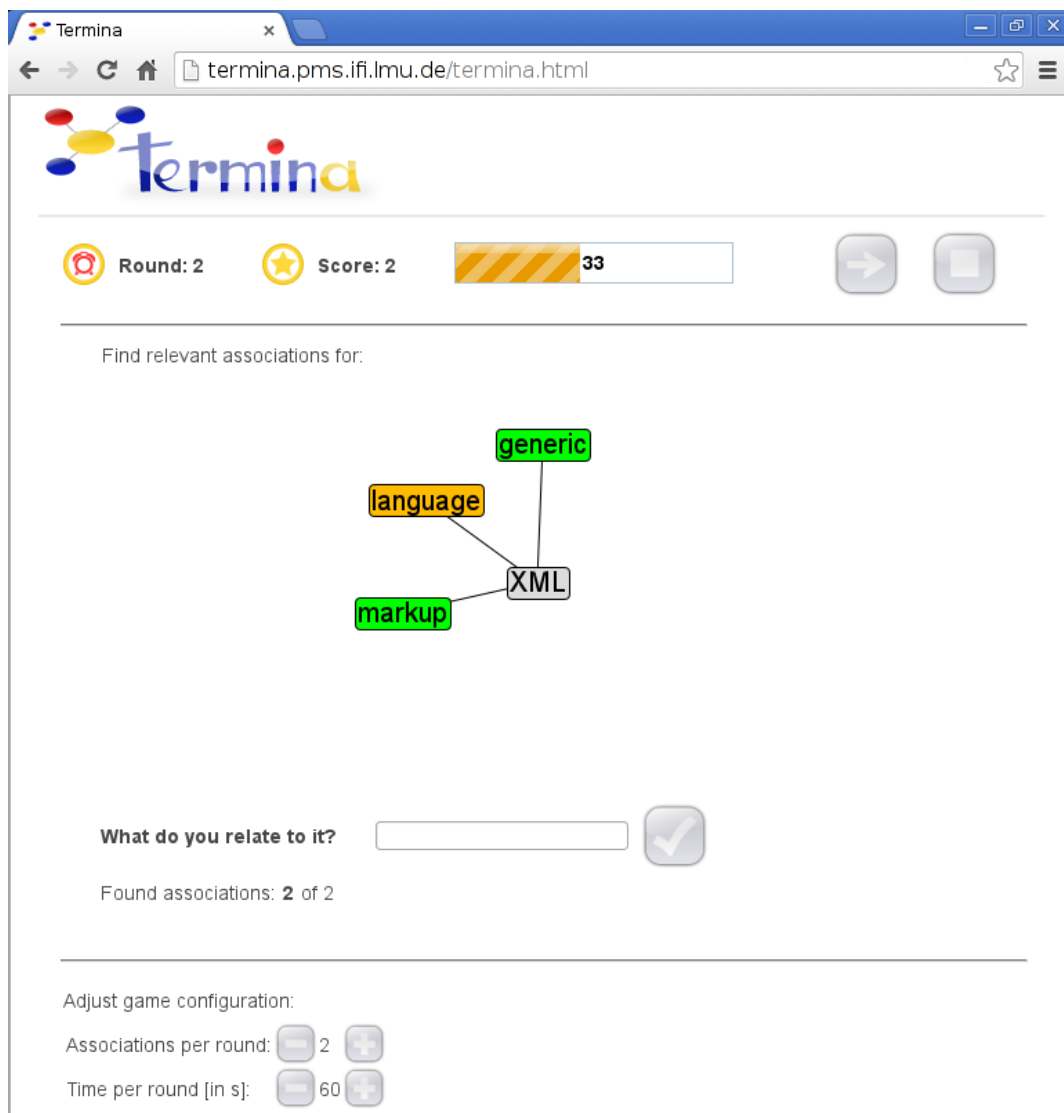
The crowdsourcing aspect is the collective generation of a relation between terms and concepts together with the strength of the relationships. In this thesis, concepts and terms are distinguished. Abstract terms for which associated terms are gathered are called concepts and such associated terms are called terms. The strength of a relationship is given by the number of students stating this associated term. The learning aspect in the setting of a university course includes the facilitation of association-based learning for students and the observation of their learning progress for both students and lecturers. Association maps are constructed automatically for the students’ and lecturers’ investigations from the gathered data. Two modes of operation are available to students: a game-like *play mode* with a time limit, adjustable difficulty, and a scoring system and a *practise mode* for training without these game elements. *Termina* is used in university courses taught by François Bry at the Ludwig-Maximilians-Universität München since 2012. Its use in other contexts for a collective generation of association maps is conceivable.

In the following sections, the principle of gathering semantically associated concepts, the different modes of operation, and the construction of association maps with *Termina* are introduced. Furthermore, motivations for students and lecturers and further application areas are explained.

## 7.1 GATHERING SEMANTIC ASSOCIATIONS

Termina can be used without any knowledge of association maps. Some knowledge of the issues of the topic suffices for participating in Termina. Students can register on Termina to save (incomplete) snapshots of their learning progress to their profile, but they can also engage in Termina anonymously. During a Termina session a term representing a concept is displayed to the student and she is asked to enter associated terms, i.e., terms related to the displayed term. A screenshot of a Termina session is displayed in [Figure 24](#).

To clarify the terms “concept” and “term”, the difference between them is that a concept is used in a more abstract sense than a term, terms may be instances of a concept, and several terms may belong to one concept. This is for example the case for different spellings, for synonyms, or for terms in



**Figure 24:** Screenshot of a Termina session in progress in which the student already stated two closely associated terms “markup” and “generic” and an unclassified associated term “language” for the concept “XML”.



different languages. In Termina, a concept has a representative name, which is displayed. Concepts are chosen by lecturers whereas terms are chosen freely by students. This rule is due to the different levels of abstraction on which concept and term are seen. Lecturers as well as students can relate terms to concepts. An initial set of associated terms is entered by lecturers (see next section), which is extended with associated terms proposed by students during Termina sessions.

After a student has given associated terms for several concepts, a summary of the terms proposed by the student is displayed, including answers of other students to show differences in knowledge. Only lecturers can classify associated terms into *close* and *far*, according to the closeness of the terms to their original concepts. Students receive the classification of terms as feedback for learning which of their proposed terms and the terms of other students represent close / far associations. Thus, they can learn from the terms other students proposed which they did not propose and complement their knowledge.

Students are asked to enter associations, and thus the gathered associated terms should be semantically related to the concept. The type of the association is unspecified, thus any type of associated terms are gathered, including more precise, broader, or opposite terms or also examples for instances of the concept. An extension of Termina for gathering the types of existing associations would make sense in order to add further semantic value and remains for future work.

The details of Termina and its two operation modes are explained in [Section 7.2](#).

#### 7.1.1 Deploying Seed Data

Termina focuses on the relation between concepts and terms and is thus suited to give an overview over a topic as well as to generate links between topics, facilitating cross-linked learning. To this aim, Termina is always based on one specific topic, in this thesis the topic of a course at university. The start page of Termina offers the possibility to switch between the available topics. Seed data, which is needed in order that students can participate, specified for Termina by lecturers. This data consists of

- concepts that are important for the topic,
- the concepts' difficulty,
- an initial set of closely associated terms for each concept, and
- concept subsets.

The concepts and closely associated terms imported by lecturers are the basis for Termina sessions. During further sessions students add more associated terms to the existing ones. The difficulty of concepts is represented as number—this thesis proposes numbers one to three—and it affects the score students can achieve in Termina sessions: the higher a concept's difficulty the higher the

achievable score. Lecturers can therefore award students more points for stating associated terms for difficult concepts than for easy concepts. A concept subset is typically given by a chapter of the course or by a subtopic, which addresses some but not all concepts of the course. Students employing Termina can then focus on learning concepts of one chapter in addition to revising all concepts of the whole course.

For generating the seed data itself several options exist. In the simplest case, which has been employed for courses so far, lecturers manually enter concepts, difficulties, and initial sets of associated terms. Here, the lecturer has full control over all data, but it takes some (limited) effort to generate the data. Usually, this is done by filling in the data into a spreadsheet in the process of going through the slides or lecturer's notes for one lecture. The data of the spreadsheet can then be imported into Termina by simply uploading the file. Another possibility for generating seed data consists in developing an algorithm which takes lecture notes or slides as input and extracts highlighted words as concepts and words on the same slide as closely associated terms. Such an algorithm is conceivable but has not been implemented as part of this thesis. Before importing seed data the lecturer should review it and optionally set the concepts' difficulties.

### 7.1.2 Classification of Associated Terms

After initial data has been set up, students can engage on Termina in one of the two modes of operation. Through their participation, associated terms are gathered which fall in two classes: associated terms which are already in the concept's set of classified associated terms, and associated terms that are not yet in the concept's set of classified associated terms. Such new terms are by default *unclassified*, waiting to be classified by the lecturer.

From the lecturers' point of view the status of associated terms is an important issue. No associated terms that are wrong should be classified as *close* because students must not be incited to learn wrong associations. Therefore, all new, unclassified associated terms entered by students are presented to lecturers for classification into either *close* or *far*. The naming of *close* and *far* resembles the fact that with associations there usually is no right and wrong because every term or concept is (more or less directly) associated with every other term or concept. Furthermore, the classification of terms for a concept is not necessarily the same in two different topics, i. e., courses. In one topic a certain term may be close to a certain concept, while in another topic it may be far. Especially homonyms and abbreviations are prone to such differences.

One might wonder why this thesis proposes the three states unclassified, close, and far for classification and not just one state, more than three states, or a numerical value. There are several reasons for this.

Having just one state means that associated terms cannot be differentiated, which makes data handling easier. However, this approach is problematic in an e-learning situation. Data that is "provided" by the lecturer needs to fulfil quality standards of being correct because students rely on the correctness.

And students may regard data available on Termina as provided by the lecturer. Furthermore, by participating in Termina students would benefit from getting to know associations of other students, but they would not be advised whether the associations are good for this topic or not. The feedback for learning thus falls short if associated terms are not differentiated.

The closeness of an association may well be classified into more than three states or a numerical value, e. g., between 0 (representing no relationship) to 1 (representing the strongest relationship, i. e., the concept itself). This would introduce a more fine-grained control of the association closeness and the feedback given to students may be richer or more precise. The problem, however, arises that the lecturer may have difficulties to judge the closeness consistently in a fine-grained manner. The mental strain for such a classification is higher, leading to more workload for the lecturer. In general, mental strain can occur if the mental classification does not match the classification provided by the system. Depending on the individual lecturer and her mental model, a different amount of states may lead to mental strain. Karasek [101] suggests to increase decision latitude to reduce mental strain. Following this approach, the lecturer could specify the amount of states herself to fit Termina's classification to her mental model. The mental strain may be reduced this way; however, the classification is no longer unified and reuse of classified terms for upcoming courses by different lecturers is difficult. Thus, the approach is not integrated in Termina but is still of interest for future work, combined with a reasonable method for reuse.

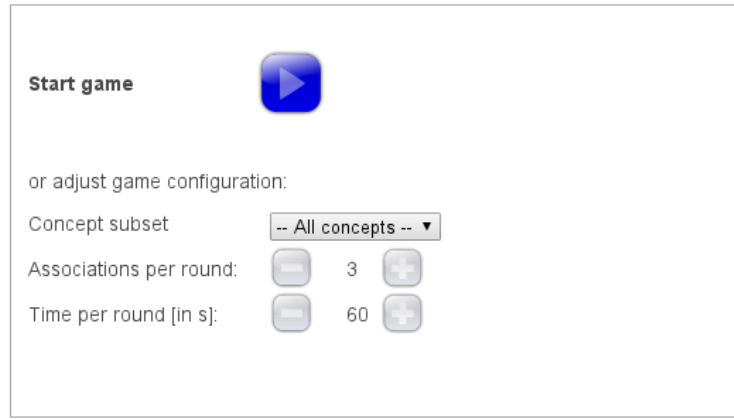
The approach to have three states on the one hand enables elaborate feedback for students and on the other hand minimises the workload for the lecturer. Therefore, this thesis proposes the approach of three states as a reasonable compromise for Termina.

## 7.2 TERMINA'S MODES OF OPERATION

After having introduced the association-based principle of Termina, the exact operation of Termina is explained. Two modes of operation are offered to students: a play mode and a practise mode. In both modes associated terms are gathered. The modes differ in the intention with which they are used by students, as described in the following.

### 7.2.1 Play Mode

Upon starting a Termina session in play mode, students can optionally adjust the default level of difficulty which consists of the time they have for each concept (60 seconds by default) and the number of associated terms to be stated for each concept (2 by default). The adjustment settings can be seen in [Figure 25](#). Students can therefore learn according to their current learning progress which might be at the beginning or already advanced. Students can



**Figure 25:** Screenshot of the page for adjusting the difficulty and selecting a concept subset at the beginning of a Termina session in play mode.

also opt to be faced with a subset of all concepts, e. g., to revise just the first chapter of the course. This option is especially helpful for courses comprising many concepts.

In Termina sessions in play mode, a normal mode alternates with a choice-based mode, which is displayed every three rounds to add some variety. In the normal mode, one concept, chosen at random within the current topic or subset, is presented to the student and she is prompted to enter terms associated to this concept for gaining points (see Figure 24 for a screenshot). In the choice-based mode, one random concept is shown together with close and far associated terms and the student has to choose the close ones in order to gain points.

For each entered (or chosen) term the student either gains points (in case it is classified as *close*), or loses points (in case it is classified as *far*), or gains no points (in case it is *unclassified*, i. e., has not yet been classified). The number of points for a term depends on the concept's difficulty as well as on the session's level of difficulty.

The following principle is followed for the scoring system: If no associated term is entered or if all associated terms entered are unclassified, the score is negative, giving the feedback that the student cannot give the relevant close associated terms for this concept and should learn more about it. As soon as half of the required number of closely associated terms have been entered, the score is zero, meaning that there already is knowledge but still not enough. For a higher number of closely associated terms, the score is positive, confirming the good knowledge of the student. Penalty points are deducted when entering far associated terms. The score therefore serves the purpose of reflecting how well students know associated terms for a concept.

This is condensed into the following formula:

**Definition 22** The *score* for one Termina round is:

$$score = (2 \cdot a_{close} - d_n) \cdot d_c \cdot d_t - a_{far}$$

where

$a_{close}$	<i>number of closely associated terms stated by the student</i>
$a_{far}$	<i>number of far associated terms stated by the student</i>
$d_c$	<i>difficulty of the concept</i>
$d_n$	<i>number of associated terms to be stated</i>
$d_t$	<i>time factor, increasing with decreasing available time</i>

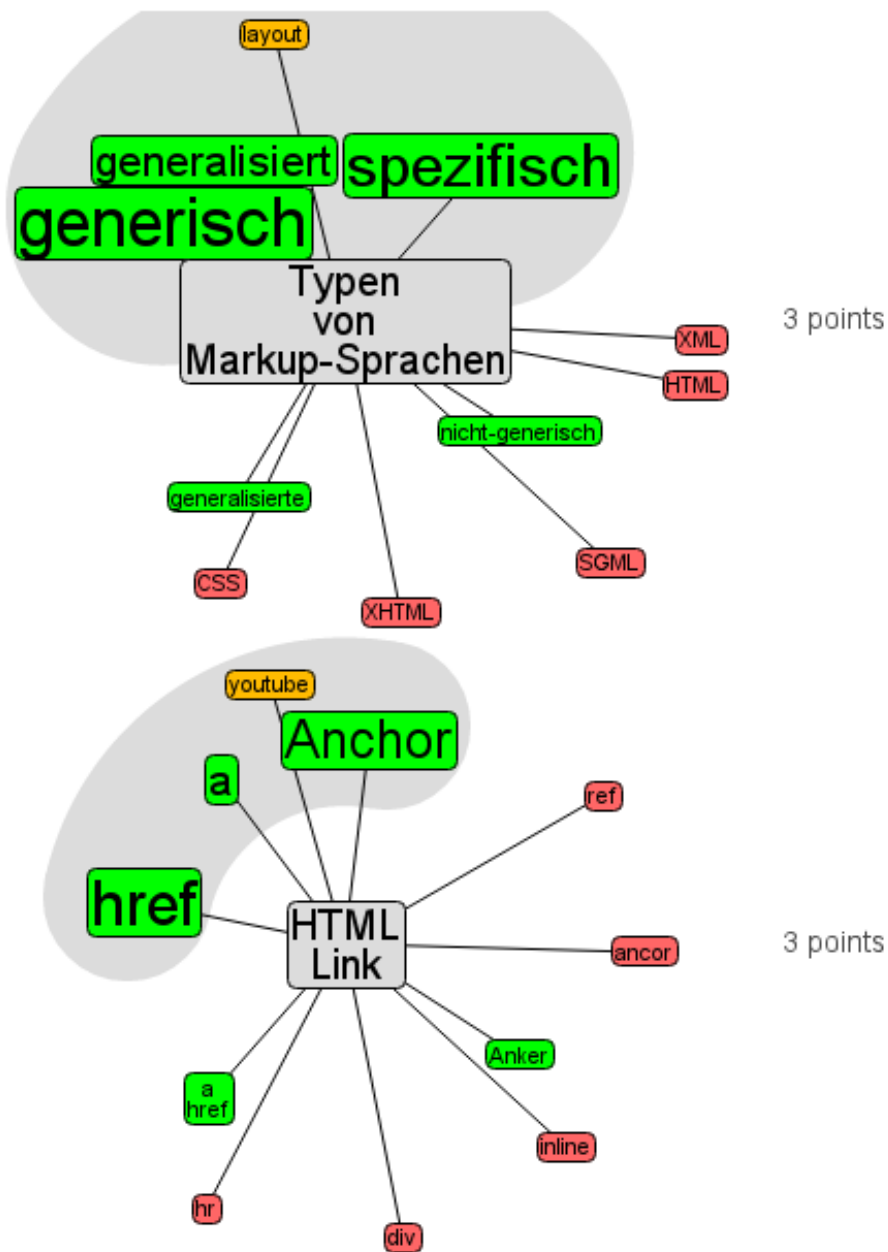
The values  $a_{close}$  and  $a_{far}$  follow from the associated terms stated by the student and their classification,  $d_c$  is defined by the lecturer, and  $d_n$  and  $d_t$  can be influenced by the student by adjusting the level of difficulty. The time factor  $d_t$  is calculated on the basis of the time the student has for stating terms associated to one concept. In the version of Termina employed in this thesis, the time  $t$  can be adjusted to 15, 30, 45, or 60 seconds and the time factor is  $d_t = 5 - t/15$ . This results in  $d_t = 1$  for 60 seconds and  $d_t = 4$  for 15 seconds. Other values might be advisable for different applications.

One round is completed and the next concept is shown either if the time is up or if as many closely associated terms as chosen in the level of difficulty, i. e.,  $d_n$ , have been entered. A Termina session finishes if associations for all concepts have been given or if the student's total score becomes negative. The latter might already happen after the first round if no close association is entered as the initial score is zero. However, after making progress over several rounds and accumulating points, students can also fail to know anything about one concept and still continue the session. For each logged in student, the total score of a Termina session is shown in a leaderboard.

At the end of the session, a summary of the concepts and associated terms is displayed in the form of association maps (see [Figure 26](#)). The layout algorithm of such association maps during and after a Termina session with a single concept in the center of the diagram is the following: Close associated terms have a green background and are positioned on a close circle around the concept. Unclassified associated terms have an orange background and are positioned on a circle with a larger diameter. And far associated terms have a red background and are positioned on an outermost circle with a yet larger diameter. The size of the associated terms increases with the number of times the associated term has been stated by students. Additionally, the student's own stated associated terms are highlighted through a grey background. Terms entered by other students are displayed as well as the ones entered by the student herself. Therefore she can monitor her progress, see gaps in her knowledge, and discover where she should learn more. She can compare her contribution to terms stated by the community of all students and learn associations of which she did not realise that they existed.

### 7.2.2 Practise Mode

The practise mode is similar to the normal mode of the play mode. One concept, which is randomly chosen, is displayed and the student enters associated terms. However, she can enter as many associated terms as she wants to. These are displayed together with their classification (close, far, unclassified) to provide



**Figure 26:** Screenshot of the summary displayed at the end of a Termina session.

direct feedback on the student's input. In contrast to the play mode, the practise mode does not impose a time constraint on the student. Thus, she can take her time to find associated terms and look up concepts in the course material. Furthermore, the student chooses when to switch to the next concept. This can be directly when seeing that she does not want to give associated terms for a concept or when she gave all terms she knows. In the practise mode, no scoring mechanism is active in order not to distract the student from practising.

The student can end the practise mode by herself or practise all available concepts. At the end, a summary of all concepts and associated terms is shown, similar to the play mode, only without scores.

The frequent presentation of different concepts and the time constraint make Termina's play mode more challenging than the practise mode. For deep learning [37], however, the practise mode is better suited because the students can focus on one concept, think about the concept, look it up in course material, and state every association which comes to their mind. In both modes of operation, elaborate feedback is given at the end of a session.

## 7.3 BENEFITS OF USING TERMINA

The feedback mechanisms with comprehensive association maps are the main incentives for using Termina both as student and as lecturer. Additionally, the association-based kind of learning provides additional benefits different from normal learning by studying lecture material and doing exercises. These different kinds of incentives, namely association-based learning, feedback for students, and feedback for lecturers, are explained in the following.

### 7.3.1 Association-Based Learning

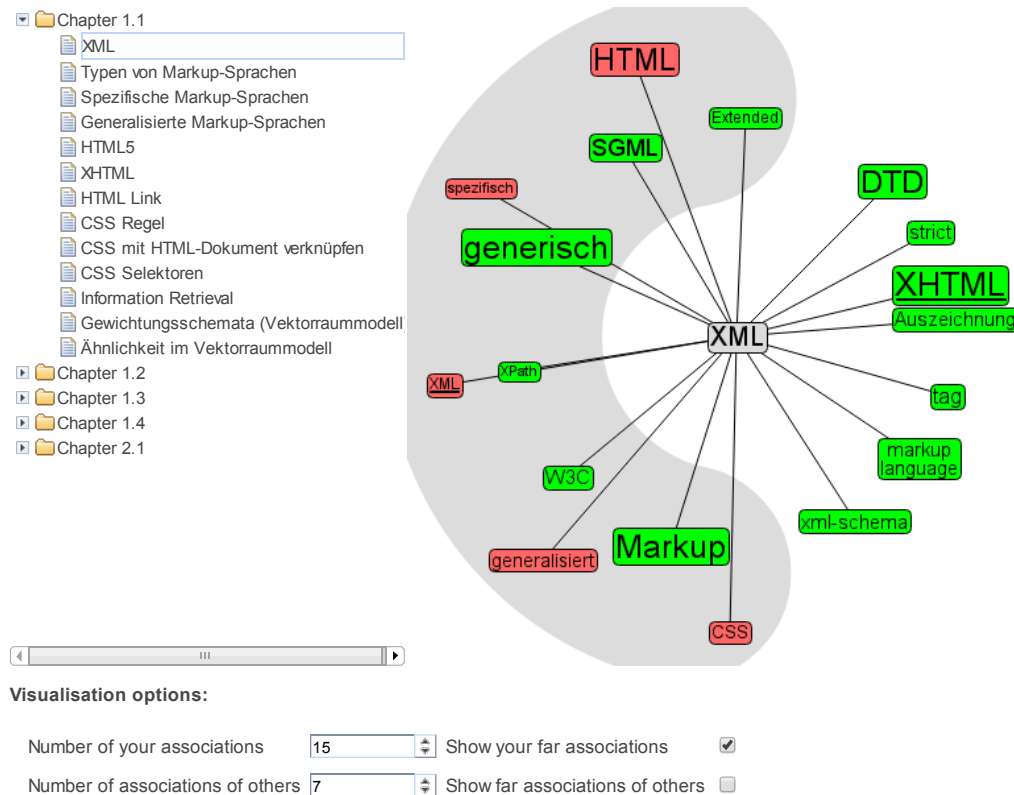
The kind of learning that can be done with Termina must be seen as supplementary to the usual learning by getting taught in class, by revising lecture material, and by completing exercises. Students learn the concepts themselves and their precise meaning the usual way, and they train their gained knowledge, test it, and advance it with Termina. In that, Termina provides variety with its association-based learning. It is furthermore challenging if the difficulty is set high enough for the student's current learning progress, and students are trained to quickly recall what a concept is about. The scoring system, the leaderboard, the time restriction during game rounds, an adjustable difficulty, and an animated user interface are gamification techniques that contribute to Termina's appeal as a game and to playing Termina being fun, especially compared to the tedious study of lecture material.

In some areas of educational psychology such association-based learning is seen as the core process how humans learn. In the behaviourist view, and especially in associationism as subordinate theory of behaviourism, knowledge is an organised collection of connections between elementary mental or behavioural units [67]. The process of learning is the acquisition of such connections. Associations, which are a kind of connections, are seen as the main part of knowledge. Shanks contrasts association-based theories with cognitive theories and concludes that associationism continues to play a major role for explaining learning [180]. Thus, also research supports the kind of learning in Termina.

## Learning progress

Control your learning progress and inspect your own associations and the associations of the others.

Green associations are already confirmed by the lecturer as being close, red associations are classified as far, and orange associations are not yet classified. Your own stated associations have a grey background.



**Figure 27:** Screenshot of the learning progress displayed on the start page of Termina for a logged in student.

### 7.3.2 Feedback for Students

In addition to the advantages, which association-based learning provides, feedback is an incentive for students to use Termina. They benefit from the following feedback measures:

- Students receive feedback for associated terms that they state and that have been classified by the lecturer. The classification as “close” or “far” on the one hand yields points to the student as a game element and on the other hand gives advice for knowing or not knowing about a concept as a teaching element.
- In the summary of a Termina session in play mode the score for each concept is shown to students. Thus, students can quickly identify which concepts have to be revised and which concepts are well known.
- Also the inclusion of associated terms stated by other students represents useful feedback. Such associations previously unknown to a student can



encourage her to study the relevant course material and to learn more about the associations. Moreover, students can compare their learning progress to the aggregated association knowledge of all students.

- For students who are registered on Termina the start page provides an aggregated view of their current learning progress (see [Figure 27](#)). They can select the individual concepts whereupon the respective associations are displayed. Also some parameters are adjustable for granular control of the displayed associations.

This individual feedback for students is important as an incentive for engaging on Termina. Also other researchers share this opinion. E. g., Bellotti, Kapralos, Lee, Moreno-Ger and Berta [11] note that for serious games, players should get immediate feedback and their contributions should get assessed. This furthermore enables the adaption of the learning mechanism to the individual participants.

Consequently, Termina as a tool that supports learning helps students with sorting terms and concepts and relating them to each other.

### 7.3.3 Feedback for Lecturers

Termina yields also feedback for lecturers. The associated terms that students enter are useful for several reasons.

- The activity of students on Termina, i. e., which concepts are played with and how many associations are stated, and its change over time can show if students may have difficulties following the course, e. g., due to time constraints, when students revise certain topics and which topics they focus on. For example, if many students (or no students) choose one specific topic, the lecturer might decide to spend additional time clarifying it.
- The associated terms gathered for one concept can reveal which concepts are well understood and which are not. Also here, the lecturer might explain a topic more closely.
- Misunderstandings about concepts can also be unveiled if many students state associated terms that are far or that are unknown but need to be classified as far. The lecturer can then steer against such misunderstanding either by classifying such associated terms as far or by telling students in class.
- Lecturers may receive new associated terms they did not think of and which are meaningful. They can use the information about the associations and include it into the next lecture or future courses. This is an example of learning analytics (see [Section 2.4.2](#)).

All of this feedback helps lecturers to learn about their students, to do better at educating students and, finally, to improve their teaching.

## 7.4 CONSTRUCTION OF ASSOCIATION MAPS

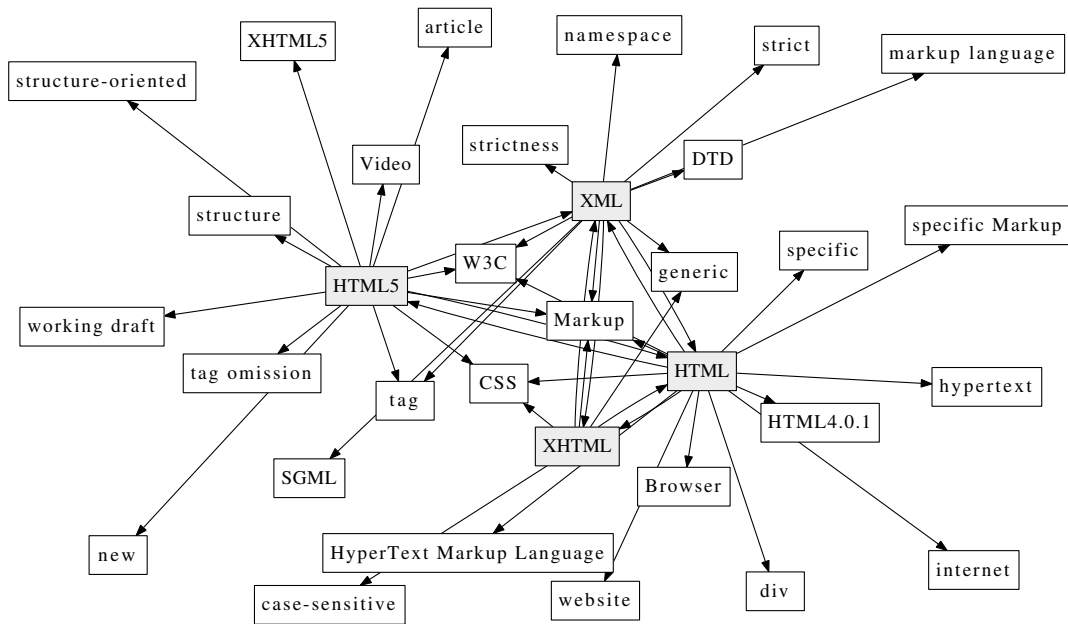
The main ideas of Termina itself have been introduced. In this section, the construction of association maps which are used for generating elaborate feedback to students and lecturers is described.

Association maps are proposed in this thesis as a simplified version of concept maps with only one proposition type, namely the association. Concept maps are designed as a visual tool for organising concepts and relating them with each other [152]. A concept map is a diagram of concepts as nodes and propositions, i. e., labelled relationships, between concepts as edges. Concept maps support active, meaningful learning and therefore represent a good way to get students to think about topics and to gain knowledge that is interconnected with existing knowledge the students have (see [Section 2.4.3](#)). Though the type of the relationships are not represented in association maps, the beneficial properties of concept maps can also be seen for association maps.

Rather than constructing concept maps directly in class or in groups as it has been suggested, e. g., in [151], this thesis proposes to generate associated terms and concepts using *Termina* and to construct one-proposition-type concept maps from the gathered data for the students' and lecturers' investigations. In the current version of Termina, association maps are only generated for one concept and its associated terms at a time. The automated construction of association maps with more concepts is intended in a future version.

In this thesis, single learners', multi learners' and combined association maps are distinguished. Single learners' association maps contain associated terms just from one learner and multi learners' association maps contain associated terms from all learners taking part in Termina sessions. Combined association maps highlight a student's associated terms in a multi learners' association map. Single learners' association maps, including close as well as unclassified and far associated terms, can help students to identify where they misunderstand a certain topic. Such association maps are employed during Termina sessions. For multi learners' association maps, only closely associations are included in order to generate "correct" association maps and also to decrease the association map's complexity. They are constructed for lecturers reviewing the progress of their students. For displaying the results after a Termina session, a combined association map is displayed which combines the display of associated terms from the student herself –these terms are highlighted in order to be easily identifiable– with the display of associated terms from other students. This incorporates the advantages of both single learners' and multi learners' association maps.

In general, a multi learners' association map cannot contain all closely associated terms that have ever been stated by students because its sheer number would overwhelm anybody looking at the association map. A good way to filter associated terms is to look at the number of times an associated term has been stated by different students, the *strength* of the association, and discard all associations with a strength lower than a certain threshold. The threshold needs to be set depending on the number of students taking part in Termina, for a



**Figure 28:** An example of an association map generated from data gathered with Termina in a course on markup languages. The concepts are represented as nodes with a grey background. The layout algorithm is based on a force-directed layout approach by Fruchterman and Reingold [62]. The association strength is not taken into account, i.e., the node distance is arbitrary.

low number of students a sensible number is 2. Associations with a strength equal to or greater than the threshold value are included in the association map, all others disregarded.

For single learners' and combined association maps, the inclusion of far associated terms can yield relevant feedback to the student. For this reason, students may optionally include or exclude such terms on demand. The threshold for the minimum strength of associated terms stated by the student obviously is one in single learners' or combined association maps, because otherwise many associated terms would not be visible. The comparison of own associations and associations by other students in combined association maps can be a valuable tool for learning.

Association maps can be generated separately for each topic subset as smaller maps or for the whole topic of a course as a big overview map. Depending on the desired level of detail different thresholds might be specified as needed.

An example for a multi learners' association map as it has been generated from data in a university course is displayed in [Figure 28](#).

## 7.5 TERMINA FOR EXPERT COMMUNITIES

The context of Termina described so far is the one of a course at university. This is the primary application area and Termina has been conceived specifically for

this purpose. However, it is not the only application area for which such a game-like e-learning tool makes sense. There are many areas that would profit from Termina.

Especially for expert communities a similar tool as Termina would make sense. In this context, expert communities are communities which are focused on one specific topic. The question and answer forum  $\text{\TeX}$  Stack Exchange [271] and the movie database IMDb [250] are examples for such communities with very different topics. The concepts could be  $\text{\TeX}$  commands and specific  $\text{\TeX}$  related issues in the first case and movies, actors, and genres in the second case. Then, participants state terms associated to the concepts. The feedback algorithm would need adjustments so that associated terms are classified collectively instead of being classified by one authority person. Then, engaging on such a version of Termina would yield new associations between the concepts, and classifications of the associations could be generated for example with a thumbs up / thumbs down system. Participants would benefit from the knowledge that is gathered this way. For example,  $\text{\TeX}$  commands could be associated with similar commands or related issues. By automatically building an association map from such data people may identify  $\text{\TeX}$  commands and issues related to each other faster. With an enriched user interface, navigating from a specific command to other related commands which provide similar functionalities or are related in another way would help finding ways to achieve a certain goal. For the movies example, a concept subset could be created in which actors for movies have to be named. Playing Termina on this concept subset would result in an application similar to a quiz. Also, similar movies or actors could be identified through Termina and leaderboards of the best participants could attract people from the community.

Deploying a tool similar to Termina in expert communities thus would be a practical choice that could lead to the generation of purposeful association maps.

In preparation of its evaluation along a university course, Termina’s design has been optimised through usability studies and iterative improvements aiming at making Termina as intuitive to use as possible [27, 116]. Results of a user study focusing on the user interface show that Termina is easily understandable and that students would appreciate Termina being used in lectures [27]<sup>1</sup>.

After these encouraging results, Termina was employed along two courses at the Ludwig-Maximilians-Universität München. The evaluation presented here is based on data gathered during the course “Programmierung und Modellierung” (in English: “Programming and modelling”), an introduction to functional programming for first year bachelor students of computer science, media informatics, and bio informatics. The course was given by Prof. François Bry. The data was gathered during the five months of the summer semester 2013 till the final exam. A small, optional survey was carried out among the participants of Termina.

For informing students about Termina, its use was suggested on the web page accompanying the course, and the lecturer announced Termina and its possibilities at the beginning of a few lectures. No rewards were offered to the students for their participation in Termina and for completion of the survey.

Test data from lecturers has been excluded for this evaluation in order to not influence the results.

In the following, the quantity of data and the quality of data is examined, results of the user study are outlined, and a brief conclusion is presented.

## 8.1 DATA QUANTITY

The total amount of data that has been gathered during the 21 weeks of the course is displayed in Table 4. To summarise, 107 concepts have been created by the lecturer, 312 participants played 6224 rounds of Termina and stated associated terms 15607 times. The number of different associated terms is 5713. So, every associated term has been stated 2.7 times for a concept on average. From these associated terms, 2496 have been stated by at least two participants, i. e., have a minimum strength of two.<sup>2</sup> From these, 1713 have been classified as close and 483 as far by the lecturer, the rest are unclassified.

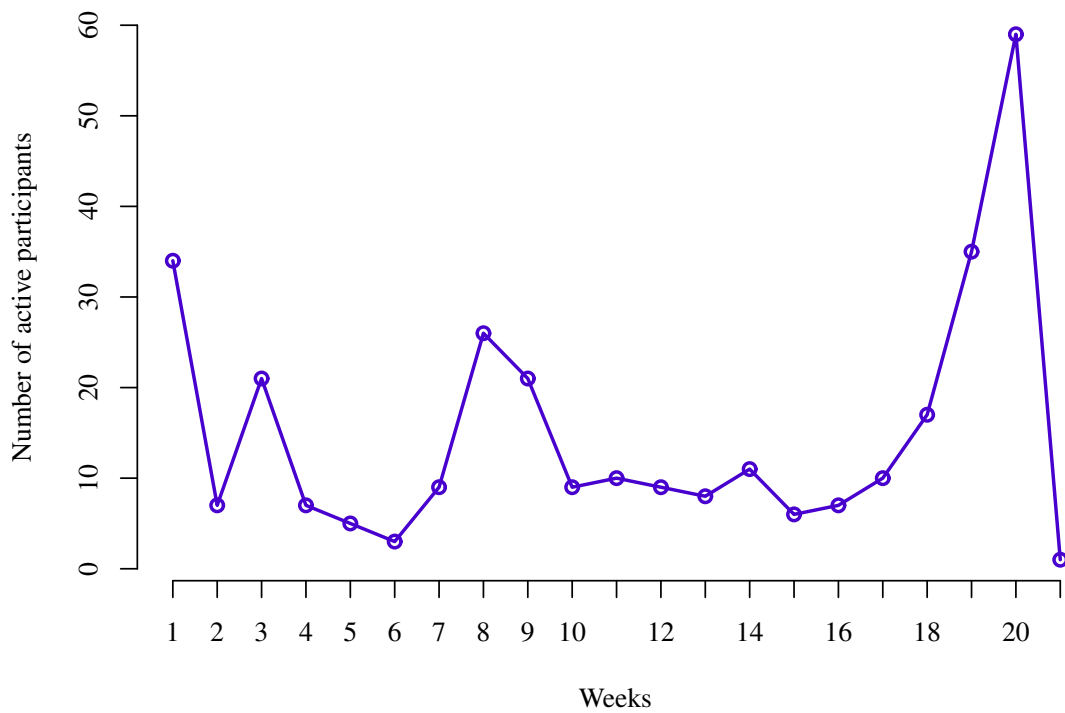
---

<sup>1</sup> The results of the usability user study are not part of this thesis because they focus on the design of intermediate prototypes and not on a practical evaluation along a university course. See [27] for detailed information on the user study.

<sup>2</sup> Only associated terms with a minimum strength of 2 are considered here because they represent the terms relevant for several students.

Participants	312
Registered participants	12
Concepts	107
Played rounds	6224
Stated associated terms	15607
Different associated terms	5713
Different associated terms with strength $\geq 2$	2496
Close associated terms with strength $\geq 2$	1713
Far associated terms with strength $\geq 2$	483

**Table 4:** Total amount of gathered data. The strength of associated terms denotes the number of times it is stated by different participants.



**Figure 29:** Number of active participants per week.

In the number of participants, not registered students who engaged on Termina several times are counted several times, due to the technical constraints of cookies. Therefore, the actual number of students who engaged on Termina is lower than 312.

The numbers show that students were engaging on Termina quite actively and that they were stating a high number of associated terms. Also the high number of associated terms per concept (53) shows that students stated a multitude of different associations that would probably not have been gathered by one person alone. It is a good sign that the number of far associated terms (483) is much lower than the number of close associated terms (1713), meaning that concepts were generally reasonably well understood.

The activity of students over time can be seen in [Figure 29](#). During the first, third, and eighth week, Termina has been promoted in class and exactly during these weeks, many students participated. During the last weeks, participation soared, probably because the students were studying for the exam and eager to benefit from playing Termina. The temporal development of the number of game rounds and stated associated terms are similar to the one in [Figure 29](#), and they can be found in [Appendix A.2.1](#).

## 8.2 DATA QUALITY

After having described the quantity of data being gathered, their quality is now assessed. For that, first the associated terms for one concept are examined, and second an association map generated from the gathered data is shown.

For the exemplary concept “map function”, 27 different associated terms have been stated by at least two students. These are listed in [Table 5](#). It can be seen that many different elements of the map function are revealed by the students: The map function is applied to lists, it is a predefined function in SML (the exemplary language for this course), it can be used in curried form [90], it is related to filter and fold, it is a higher-order function and it takes a unary function as parameter. Most associated terms (23) are classified as close by the lecturer, four are classified as far, and none are unclassified. It follows that the stated associated terms are quite accurate and give a good overview over the concept of the map function.

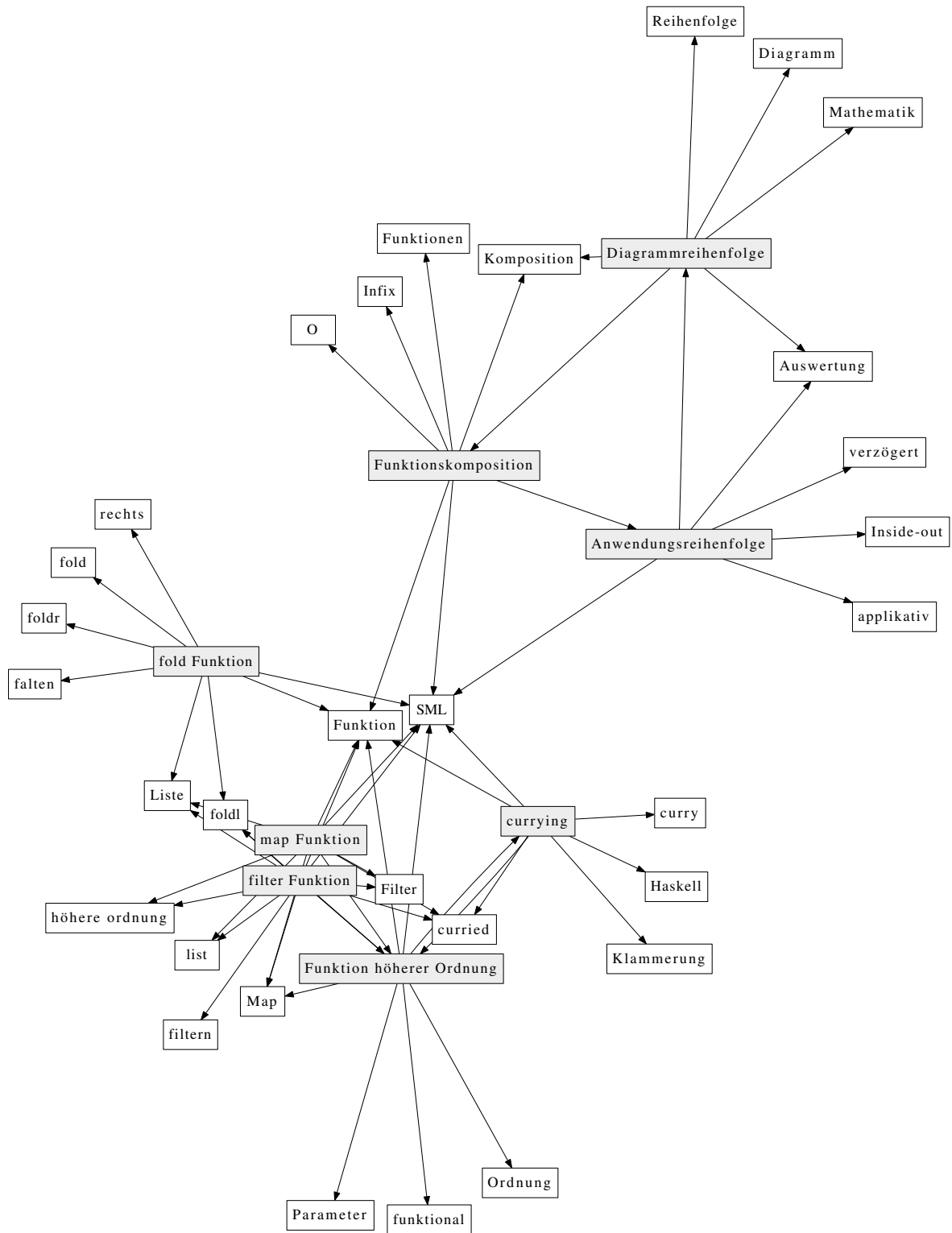
An association map of a concept subset of the course (a chapter about higher-order functions in week 8) generated from concepts and gathered associated tags is depicted in [Figure 30](#). The concepts are “Funktionskomposition” (function composition), “Anwendungsreihenfolge” (application order), “Diagrammreihenfolge” (diagram order), “Funktion höherer Ordnung” (higher-order function), “map Funktion” (map function), “fold Funktion” (fold function), “filter Funktion” (filter function), and “currying”. All other nodes in the association map are close associated terms that have been stated by at least four students. Some such terms are concepts themselves. Therefore, the concepts are interconnected, either directly or indirectly over other terms.

close	far	unclassified
list (33)	O (4)	-
function (12)	associative (3)	
SML (10)	recursion (2)	
curried (6)	together (2)	
filter (6)		
higher-order (5)		
map (5)		
higher-order function (4)		
curry (3)		
fold (3)		
function application (3)		
higher function (3)		
unary function as parameter (3)		
all list elements (2)		
applying (2)		
element (2)		
foldl (2)		
foldr (2)		
iteration (2)		
list of values (2)		
parameter (2)		
polymorphism (2)		
unary function (2)		

**Table 5:** Associated terms, with their association strength in brackets, for the concept “map function” which have a minimum strength of two. The terms have been translated into English for better understanding, the original terms can be found in Appendix A.2.1.

When the lecturer first saw this association map, he immediately noticed two associations that caught his interest. The first association is from “currying” (creating chains of functions, see, e.g., [90]) to “Klammerung” (in English: “parenthesizing”). This association was unusual on his first sight but valuable on the second sight because how and which parentheses are set influences the function calls. He considers students who give this association as knowing what currying is about, as they probably tried currying in practise. The second association is “Funktionskomposition” (in English “functional composition”) to “Anwendungsreihenfolge” (in English “application order”). The lecturer sees the association ambivalently: One meaning of application order is that of the order in which an expression is evaluated. In this strictly functional point of view the association from functional composition does not make sense. Another meaning of application order is one of two possibilities of functional composition in a broader sense. In this point of view the association is sensible. From looking at this association map for a few minutes, the lecturer could see this issue of ambiguity which he could address in the next lecture. Thus, immediate benefits emerge from the lecturer’s inspection of the association





**Figure 30:** Association map of one concept subset of data gathered during the course. Associations were filtered to only include close associations with a minimum strength of four.

map displayed here. Also for students it can be useful for learning and revising the topic.

This association map has been generated with the GraphViz [245] software. The generation of maps with multiple concepts is not included in the current version of Termina.

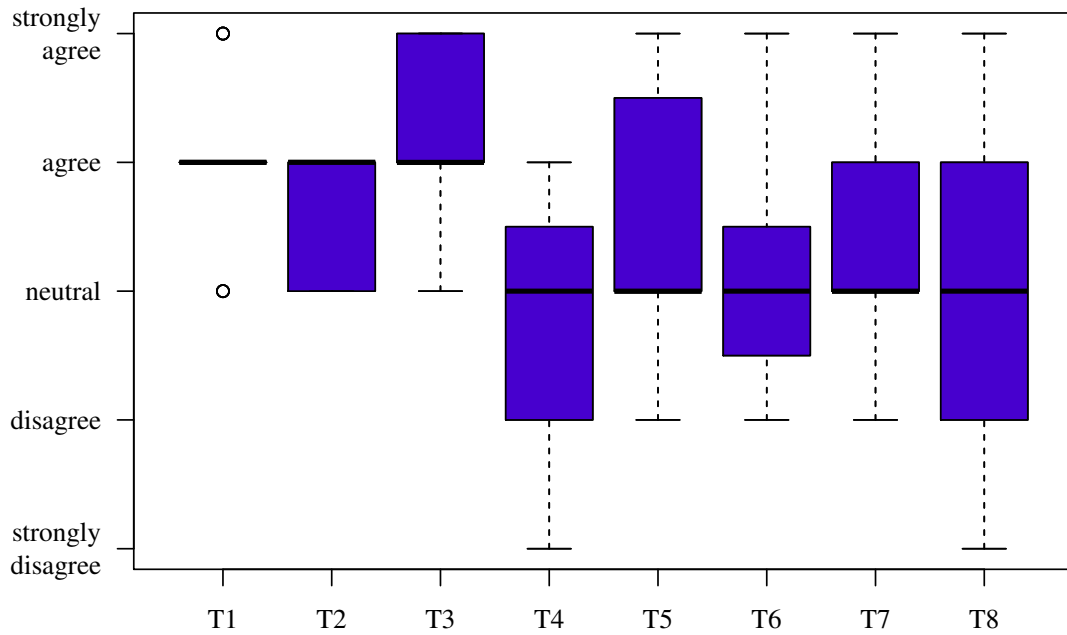
### 8.3 USER SURVEY

After playing Termina, students were asked to complete a survey. The survey consists of eight questions asking for their agreement on a certain statement and one question requesting a free form comment. The choices for the agreement on the statements represent a 5-point Likert scale ranging from “strongly disagree” to “strongly agree”.

11 students completed the survey. The complete dataset of the questions and answers can be found in Appendix A.2.2. Figure 31 displays the results of the survey. The results for first three questions are statistically significant (at  $p < .05$ ), the others are not. For measuring significance, a two-sided one-sample t-test is used to test against the null hypothesis of a mean of 3 (neutral) [239]. The normality assumption required for the t-test is confirmed with the Shapiro-Wilk test [181]. The survey shows that Termina is intuitive to operate (T1, t-test: mean=4,  $t=5.2$ ,  $p=.0003$ , Shapiro-Wilk:  $p=.008$ ), that the possibility that students can define the difficulty themselves is good (T2, t-test: mean=3.6,  $t=4.2$ ,  $p=.002$ , Shapiro-Wilk:  $p=.00005$ ), and that it makes sense to display the other students’ associations (T3, t-test: mean=4.3,  $t=5.4$ ,  $p=.0003$ , Shapiro-Wilk:  $p=.009$ ). The result for question T5, i. e., that the summary is helpful for learning, is not statistically significant at  $p < .05$  but only at  $p < .1$  (t-test: mean=3.5,  $t=1.6$ ,  $p=.1$ , Shapiro-Wilk:  $p=.09$ ): it shows a tendency that students receive it as helpful. The result for question T7, i. e., that Termina encourages students to look up terms in the lecture notes, shows a weak positive tendency, however, not even at  $p < .1$  (T7, t-test: mean=3.3,  $t=1.2$ ,  $p=.3$ , Shapiro-Wilk:  $p=.1$ ), thus the result has to be dismissed. The other questions show no tendency at all.

The free form question yielded several responses which are summarised in the following (see Appendix A.2.2 for the original responses in German).

- The visualisation of the terms is very good.
- The game interface does not fit onto one student’s screen.
- More associated terms should be added as seed terms.
- One associated term has not been classified correctly according to a student.
- Creating a learning game for a course is very good. However, learning with Termina alone is not sufficient, because the knowledge that two terms are related is sufficient and the knowledge about a term is not needed.



Item	Question
T1	Termina is intuitive to operate.
T2	The possibility to define the difficulty for oneself is good.
T3	The display of other players' associations is sensible.
T4	The total score shows how well one knows the topic.
T5	The summary of terms and associations is helpful for learning.
T6	The summary of terms and associations shows how well one knows the topic.
T7	The Termina game encourages me to look up terms in the lecture notes.
T8	I learned new connections through the summary.

**Figure 31:** Results of the Termina survey. Boxplots of the answers are displayed with median, upper and lower quartiles, range of answers, and outliers.

- The game is funny.

After the survey, the game interface has been changed for fitting better onto small screens. And also the one associated term has been re-classified. The response that learning with Termina is just one part of learning shows that the aim of Termina has been understood correctly by students.

## 8.4 CONCLUSIONS

The evaluation shows that students are motivated engaging in the game-like e-learning tool Termina and that it can be a useful tool to support students with learning. Students are aware of its limits of supporting learning which is good.

By employing Termina in several consecutive courses, a valuable dataset can be established for future courses and for the generation of association maps. Thus, Termina is employed the second time for the same course on web information systems during the winter semester 2013/2014. The benefits are that students can play with a bigger dataset of classified terms and that lecturers have less work with the classification of terms compared to the generation of seed data.

One longer term goal of Termina is a contribution towards collecting data for learning analytics, the collection and analysis of data about learners for improving learning [182]. The analysis of the data can have impact on the way of teaching classes and also on the way students learn. To this aim, further, instant feedback to the lecturer is needed and Termina should be in use for several consecutive courses.

Part IV

## SOFTWARE ARCHITECTURE



## 9 | SOFTWARE ARCHITECTURE

The platforms metropolitalia and Termina together with their applications Mercato Linguistico, Poker Parole, and the two modes of operation of Termina have been implemented with the same software architecture. They share common source code and each application includes adjustments of the common implementation. The platforms even share the code basis with two other platforms, ARTigo [233] and Accenti Urbani [228], and their different applications. This modular concept has the advantage that other platforms can directly benefit from improvements implemented for one of the platforms. For managing the complexity arising from this conceptual design a specific structure has been conceived. Here, the JBoss Seam Framework [266] helps with the organisation.

For the implementation of the platforms, the following tools are employed:

- Seam Framework: basis for application development
- JBoss Application Server: Web server
- PostgreSQL: relational database
- Apache Solr: search platform

In the following sections, the Seam Framework with its most important mechanisms is introduced and the need for an application server is explained. Furthermore, the project's modular concept is presented and its database structure is outlined. And finally, search with Solr is briefly explained.

### 9.1 SEAM FRAMEWORK AS BASIS

The Seam Framework (short: Seam) has been developed by JBoss for easing the implementation of applications making use of Java Platform Enterprise Edition 5.0 (short: Java EE 5) technologies. As a framework, it provides a basis for implementing mainly Web applications. It supports the implementation regarding several aspects, which are described in this section. Further information can be found on the website [266] or documentation books, e. g., [223]. These sources also serve as references for this section.

#### 9.1.1 Foundations on Java EE 5

Java EE 5 is a standard for the Java Platform that was released in May 2006 [272] for building enterprise Java applications and that includes, amongst others, the two important technologies Enterprise JavaBeans 3.0 (short: EJB3)

and JavaServer Faces 1.2 (short: JSF) [268]. EJB3 is a “lightweight framework based on Plain Old Java Objects (short: POJO) for business services and database persistence” [223]. A POJO is a usual Java object with few dependencies on other objects. Using EJB3, components can be created that access the database, perform computations on the fetched data, write back changes to the database, and offer interfaces for being accessed by other components of the software. JSF is a technology for Web development that offers graphical components for constructing web pages and supports creating templates for a consistent layout. The two technologies are complementary to each other and each is well designed on its own. However, they do not work together very well, as the components are not aware of each other [223]. This makes it difficult to develop applications integrating both standards because additional, redundant work has to be done, leading to more code, and possibly more errors. Seam integrates both technologies and provides a well-designed structure of components helping with application development.

The newer Java EE versions 6 and 7 integrate many of the useful components Seam provides. This shows that Seam was leading the way of Java Web development. The development of newer versions of Seam as a complete framework has been halted as of July 2013. However, the development of individual projects focusing on parts of Seam is continued [266].

### 9.1.2 Extended Persistence Context

With EJB3, the persistence context, i. e., access to the database, is by default limited to the method call. This means that if a JSF component calls a method retrieving an object from the database, the persistence context is closed after the method is completed. If the JSF component then wants to access an attribute of the object which has not yet been loaded –for example, an object from a one-to-one relationship in the database–, it fails. As a workaround, the method must retrieve the relevant attributes from the database before returning the object to the JSF component. While this works, it is cumbersome.

Seam therefore offers an extended persistence context: The persistence context is held open until the web page is completely generated, offering the possibility to retrieve attributes on demand – which is called lazy loading.

### 9.1.3 Context and State Management

Seam includes the notion of contexts maintaining instances of objects –thus maintaining state– and having different lifespans. The contexts Seam provides are, ordered by descending lifespan, the application, the business process, the session, the conversation, the page, the event, and the stateless context. Each context has a defined lifespan, e. g., the application context exists while the application is running and the page context exists while the user is on the same web page. The conversation context can be started and ended as needed. It is a powerful context which can be used for maintaining states over several



web pages. Also, several conversation contexts can be started and the active context can be switched between them. In each context, instances of objects are maintained. For example, the data of a currently logged in user is usually stored in the session context for being available as long as the user is on the website. In the event context, information could be stored that is important only for the instant a web page is generated.

One further addition to context and state management is the model of pageflows. A pageflow is a defined operation sequence of pages and actions which is defined through a finite state automaton with the pages as states and actions, i. e., method calls, or decisions as transitions between states. With pageflows, complex work flows can be modelled. Each pageflow is wrapped in a conversation context for maintaining its state.

#### 9.1.4 Dependency Bijection

Adding to context management, Seam is based on so-called dependency bijection, i. e., dependency injection and dependency outjection. If a Java object depends on another object, it must have some reference to an instance of the other object. For getting an appropriate instance, all that has to be done is stating that the other object should be given to the object by the framework, i. e., “injected”. In Seam, this is achieved by the Java annotation `@In`. The framework looks whether an instance of the desired object exists in the current context –or contexts of a longer lifespan– and returns it, or it may create a new instance if required. Injection thus support the developer by managing instances of objects, instantiating them upon their usage, and disposing them if they are no longer needed.

Similarly, object instances can be outjected into a context so that they can be injected into other object instances depending on them. Outjection thus represents a storage of an object instance in a defined context, and injection represents its retrieval.

#### 9.1.5 Automated Testing

Especially in agile development processes, support for automated testing of individual components of a project (so-called “unit tests”) as well as of the interaction of components (so-called “integration tests”) is important [48]. Automated tests can increase the probability that changes in the source code do not break functionalities that have been working before. Seam facilitates unit and integration testing by supporting existing test suites like JUnit and TestNG. Seam furthermore provides helper methods for easing complex integration tests that depend on the whole Seam Framework and Java EE techniques.

### 9.1.6 Web 2.0 Applications with Ajax

Ajax, which stands for “Asynchronous JavaScript and XML”, is a widely employed technique for generating Web 2.0 applications [223]. Ajax adds interactivity to web pages without triggering a reload of the whole web page. Thus, interactive web applications can be built using Ajax in which the server responds to input supplied by the user. This is important for GWAPs that live from entertaining interactivity and also for the interactive applications of *metropolitalia* and *Termina*.

Seam supports Ajax directly through components for web pages which include the whole client-server communication necessary for Ajax. Thus, developers do not have to write JavaScript and XML in order to benefit from Ajax techniques, but they can merely use the supplied Ajax components.

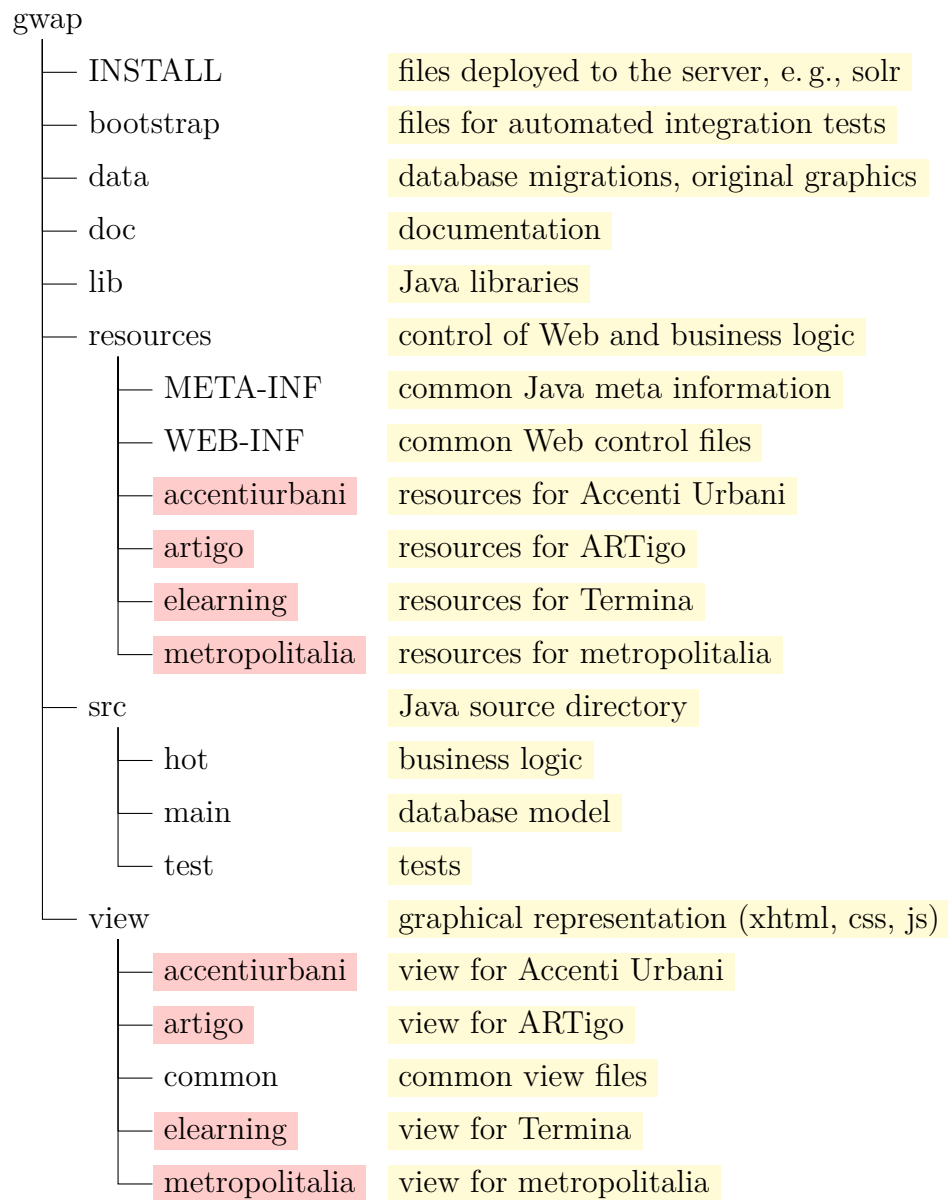
Increased use of Ajax usually leads to an increased load of the database because more requests for data have to be processed by the server. Making use of Seam’s extended persistence context and Seam’s context management, the load can be absorbed mostly by the web server [223]. Thus, Seam readily supports a high amount of Ajax requests.

All of these techniques and mechanisms integrated in the Seam Framework facilitate the implementation of interactive Web platforms, like *metropolitalia* or *Termina*, through abstract, high level code. The Seam Framework has been chosen instead of other modern Web frameworks, like Ruby on Rails or Grails for several reasons: Seam is written in the common programming language Java, it is based on the widely acknowledged Java EE standard, it promotes a well structured source code, and it is scalable on several application servers. Furthermore, the generic approach of Java EE promotes a generic software architecture and a genericity concerning the contents, which is ideal for *Agora*’s generic model.

## 9.2 APPLICATION SERVER

Applications developed with the Seam Framework have to be “deployed” to a dedicated server running the applications. The deployment is the process of copying an archive of relevant files (usually including (x)html files, compiled Java classes, and meta information for configuration) to the application server. An archive can be an **enterprise archive** (EAR), as it is done for *Termina* and *metropolitalia*, or simpler **Web application archives** (WAR), which do not contain enterprise mechanisms like a persistence context. An EAR can include one or more WARs containing (x)html, css, and javascript files and **Java archives** (JARs) containing compiled Java classes.

One freely available, open source Java application server, that is used for *Termina* and *metropolitalia*, is the JBoss Application Server (JBoss AS) [253]. JBoss AS has been chosen because, first, JBoss AS can run in a cluster, i.e., as a connected group of servers, for running applications on all servers for



**Figure 32:** Directory tree of the “gwap” project which contains the source code for the platforms accentiurbani, artigo, elearning (Termina), and metropolitalia. The highlighted directories are specific for the platform, all other directories contain common files.

load balancing. This functionality is used for the metropolitalia platform. It furthermore makes the platform more fault tolerant because one server may fail without service interruption of the platform. And second, JBoss AS fully supports Seam as both are from the same company. More information about JBoss AS can be found in [96].

### 9.3 MODULAR CONCEPT

Given the framework that Seam provides, it is not immediately apparent how to structure a project containing the implementations of applications for several platforms. Some adjustments to the build system and the directory structure were necessary to cope with the multi-platform project.

Figure 32 shows the directory tree of the main project, called “gwap”. There are many directories that contain common code, which are the default ones for a Seam project, and there are directories specific for each platform. For generating an archive containing all files necessary for one platform, the files from the common directories are included first and they are then overwritten and supplemented by platform-specific files in the platform-specific directory. For example, for building the Web archive for the platform `metropolitania`, the shared files for generating the `view`, i.e., the website, are included from the subdirectory `common`, and then files from the subdirectory `metropolitania` are added, possibly overwriting existing files in the archive.

This structure helps to keep the platform-specific files separated, which are mainly files for controlling the graphical interfaces of the web pages, i.e., the folders `resources` and `view`. And it keeps files that can be reused for other platforms together, most importantly, Java files containing the database model, business logic, and tests (subdirectories of `src`).

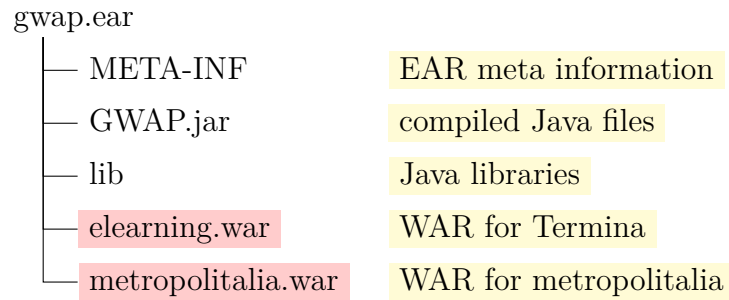
The structure makes it easy to add new platforms to the project. The following steps have to be performed:

- create new directories in the “resources” and “view” directories,
- copy the contents of the `WEB-INF` folder of an existing platform (e.g., `resources/metropolitania/WEB-INF`) to the directory created before in `resources`, and
- edit the copied files so that they contain the correct platform name.

The file `build.properties` controls which platforms are enabled for deployment to the server. Several platforms can be enabled at once, resulting in several WARs being generated and included in the EAR. The structure of the generated EAR can be as shown in Figure 33.

### 9.4 DATABASE STRUCTURE

The database which is used by the platforms is the relational open source database PostgreSQL [265]. Relational databases are well supported by Seam, and PostgreSQL, together with MySQL, is one of the most advanced and freely available relational databases [187]. The choice of PostgreSQL instead of MySQL was driven by the Play4Science members’ familiarity with PostgreSQL and by its high performance in previous applications. The database could be replaced by MySQL or other relational databases supported by Seam and Hibernate [247] with only small modifications.



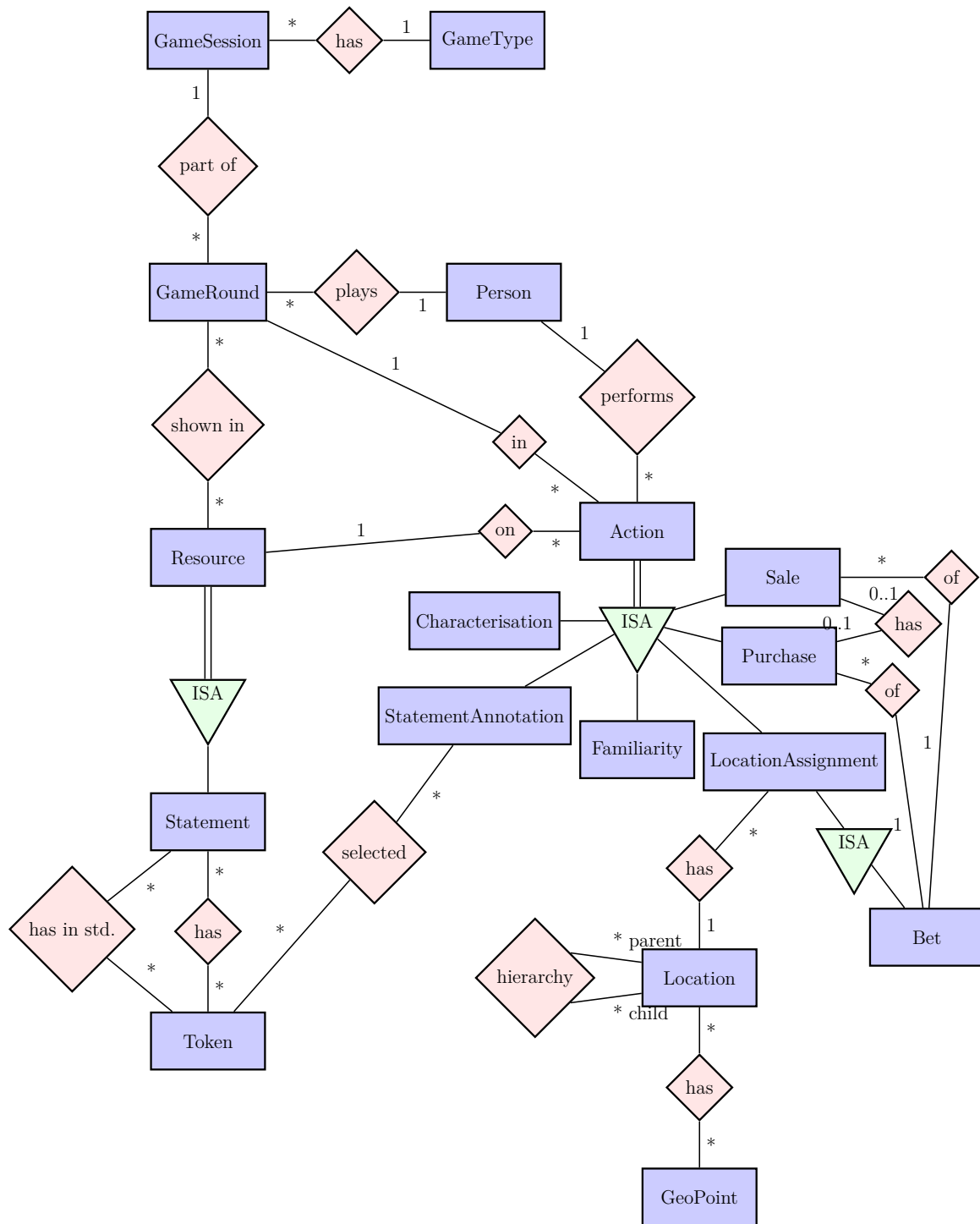
**Figure 33:** Important files and directories of the gwap EAR. This exemplary EAR contains the files deployed to an application server for the platforms elearning (Termina) and metropolitalia.

Agora name	Database name
Assessment	Bet
Geographical characterisation	LocationAssignment
Number with relations	Hierarchy
Phrase	Statement
Word selection	StatementAnnotation
Social characterisation	Characterisation
Symbolic good	Resource
Unveiling of meaning	Familiarity
User	Person

**Table 6:** Differences in naming between Agora’s data schema and the database.

The database structure is automatically generated from the definitions in the Java classes (in the folder `src/main`) by Hibernate [247]. Annotations are used to specify the types of relationships between entities and what data type is used for storage of attributes. An extract of the database tables important for the platforms metropolitalia and Termina is depicted as an entity-relationship diagram in Figure 34.

Although only an extract is shown, the diagram already has many entities and relationships. The entities mostly, but not always, have the same name as in the data schema of Agora (Figure 3 on page 33). This is partly due to forbidden names in PostgreSQL like “User” (instead, “Person” is used) and partly because of historic names from previous development versions. The main differences are outlined in Table 6. An additional minor difference to Agora’s data schema is that the dividend is directly stored in the corresponding action instead of it being a separate entity. But except these smaller differences, the diagram represents Agora’s data schema. Thus, it can be seen that Agora’s data structure can be implemented very well in a relational database.



**Figure 34:** Entity-relationship diagram in Chen’s notation [31] of a part of the database relevant for metropolitalia. Rectangles represent entities and diamonds represent relationships between entities. Inverted triangles represent an “is-a” subclass relationship. Note that for simplicity, only one relationship between Action and Resource is displayed instead of four relationships from LocationAssignment, Familiarity, StatementAnnotation, and Characterization (but not Sale and Purchase) to Resource.

## 9.5 SEARCH WITH SOLR

One functionality which is important for metropolitalia’s public visibility is its search for phrases. In order to provide a fast and flexible search, Apache Solr [232] is integrated. Solr is an open source enterprise search server based on the Apache Lucene search engine [186]. Solr is deployed to an application server like JBoss AS, it responds to query requests via HTTP, it performs extensive caching, it can be clustered on several servers, it can import data from relational databases, and it has a web interface for administration and configuration.

For metropolitalia, Solr is configured to regularly import data from the database. A full import was performed once when metropolitalia went public. Since then, Solr imports data that are new or that have changed since the last import. Such incremental updates of the search index are fast and do not use much processing power, so they can be performed every five minutes. The imported data include:

- the phrases in an Italian language variety,
- their meaning in standard Italian,
- the selected relevant words,
- the region names of the geographical characterisations, and
- the social characterisations for age, gender, and level of education as numbers.

The data are imported into separate fields of the Solr search index. A search query is then transformed from the user’s input to a Solr query. Simple search terms are forwarded directly to Solr, for example, the search for “Roma” (the capital of Italy and its region) yields phrases containing “Roma” in any of the following Solr fields: in the phrase, in its meaning, in the selected relevant words, and in the region names of its geographical characterisations. The social characterisations in contrast have to be transformed to the Solr query language. For example, “giovane” (in English “young”) is transformed into a Solr search for “1” in the gender field, as social characterisations are stored as numbers and 1 is the representation for “young”. Then, the search results are phrases for which users have characterised their speakers as young.

Besides the retrieval of search results, Solr assists with ranking the phrases according to their relevance to the query by taking multiplicities of characterisations and selections of relevant words into account. For example, a search for a certain word positions phrases for which the word has been selected as being relevant in earlier positions than others. Characterisations are treated similarly: For a search for “young”, phrases for which many users have characterised their speakers as young are positioned in front of other phrases.

To sum up, Solr allows flexible search queries, ranks the results sensibly, returns them quickly, and has a low impact on the database load.





Part V

CONCLUSION



This thesis examined how crowdsourcing techniques can be applied for empirical research in sciences oriented on humans –focusing on linguistic field research– and for e-learning.

The market-based, generic operating system Agora has been conceived for gathering data in crowdsourcing applications. Assessments of symbolic goods with their estimated agreement proportions provide additional data on properties of symbolic goods, which can support studies on the perception of languages. Agora is used for building two applications for Italian linguistic field research. Mercato Linguistico aims at gathering data on phrases having widely known linguistic properties whereas Poker Parole focuses on phrases having relatively unknown characteristics. The two applications collect complementary data. Both applications have been deployed on the platform metropolitalia. The evaluation of data collected from August 2012 until October 2013 yields several preliminary results: First, Agora’s concept of assessments with their estimated agreement proportions is widely accepted by users. Second, users are relatively good at estimating this proportion as more than half of all estimations deviate at most 20 percentage points from the calculated agreement proportion. Third, Agora can provide more precise distributions than usual characterisations. These three results imply that crowdsourcing can indeed produce research results of good quality for empirical research. And finally, also an expert in the area of variety linguistics on the Italian language sees metropolitalia as a promising step towards new methods of gathering data.

The game-like e-learning tool Termina has been conceived in the context of university courses for supporting students with association-based learning and lecturers with teaching. Lecturers define a set of concepts and students playfully state associated terms for these concepts. Then, lecturers classify the terms into close and far. From the collected data association maps are constructed, which represent an abstract comprehension of topics. The evaluation shows that Termina is well-received by students, that the principle of collaborative learning is important for students, and that lecturers can directly benefit from looking at association maps by recognising students’ misconceptions.

To summarise, the application of crowdsourcing techniques can be seen as successful in both areas. Of course, there is room for improvements and extensions. Therefore, some issues which should be addressed in future work are outlined in the following.

LINGUISTIC EVALUATION OF DATA GATHERED ON METROPOLITALIA The data gathered by the end of October 2013 can provide first insights into the

richness of data and meta-data contributed by the users, as it is done in the evaluation presented in this thesis. The next step consists in an extensive evaluation on a larger dataset and, furthermore, from experts in variety linguistics on the Italian language. Therefore, the popularity of metropolitalia has to be increased so that more users are attracted and a higher amount of data is gathered. To this aim, more personal contacts in Italy should be approached for a wider reach, the blog accompanying the platform should be actively maintained, the platform should be present in social networks, and a cooperation with an Italian research institute may be sought. This is an ongoing process that takes some time and that is currently in progress. After having collected more data, a thorough linguistic evaluation should be undertaken: Besides an evaluation of single phrases and their characteristics as presented in this thesis, a higher percentage of all phrases collected needs to be evaluated. And a comparison of data collected on metropolitalia with data collected by traditional means could further illustrate the validity of crowdsourced research results.

**APPLICATIONS BUILT WITH AGORA FOR OTHER RESEARCH AREAS** As Agora has been conceived as a *generic* operating system, it can be employed for building crowdsourcing applications in other research areas. This property of Agora should be exploited, for example, as suggested in [Section 5.6](#) for gathering perception data of artworks in art history. Building applications for art history with Agora is possible with low effort because images of artworks can be used as symbolic goods and art epochs and artists can be their characteristics.

**COLLECTION OF LABELLED RELATIONSHIPS FOR ASSOCIATION MAPS** So far, with the data gathered with Termina, association maps are constructed having one type of relationship, i. e., the association type. Based on these relationships, students could be prompted to specify their type in an additional e-learning application on the Termina platform. This extension would enhance association maps to concept maps for giving an enhanced overview of a certain topic.

**SEMI-AUTOMATIC CLASSIFICATION OF ASSOCIATED TERMS IN TERMINA** An aspect that can be time-consuming for lecturers in Termina is the classification of associated terms stated by students. Here, students could assist lecturers by voting on associated terms of other students after each Termina session, for example, through a thumbs-up / thumbs-down vote. The aggregated votes can be displayed as suggestions for the lecturer for quicker judgements and more efficient classification.

For research in the area of this thesis, no absolute validity can be achieved. A critical look on the research results of crowdsourcing for empirical research reveals that the users' acceptance of Agora's assessments is strongly supported by the evaluation, the good performance of users at estimating the agreement proportion is also supported, and the precision of the distributions is supported in exemplary cases, yet needs further evaluation on a larger dataset. Regarding the results for crowdsourcing in e-learning, Termina's acceptance by students

is strongly supported by the research results and its benefits for students and lecturers are supported, but would profit from further evaluations.

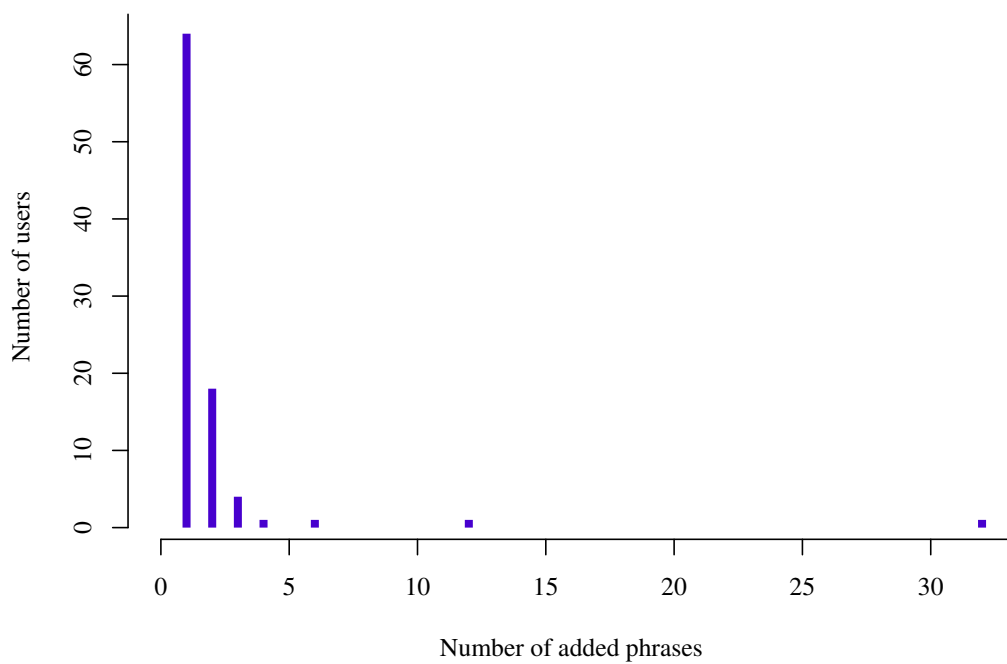
In conclusion, promising results have been achieved that provide a solid foundation for future research.



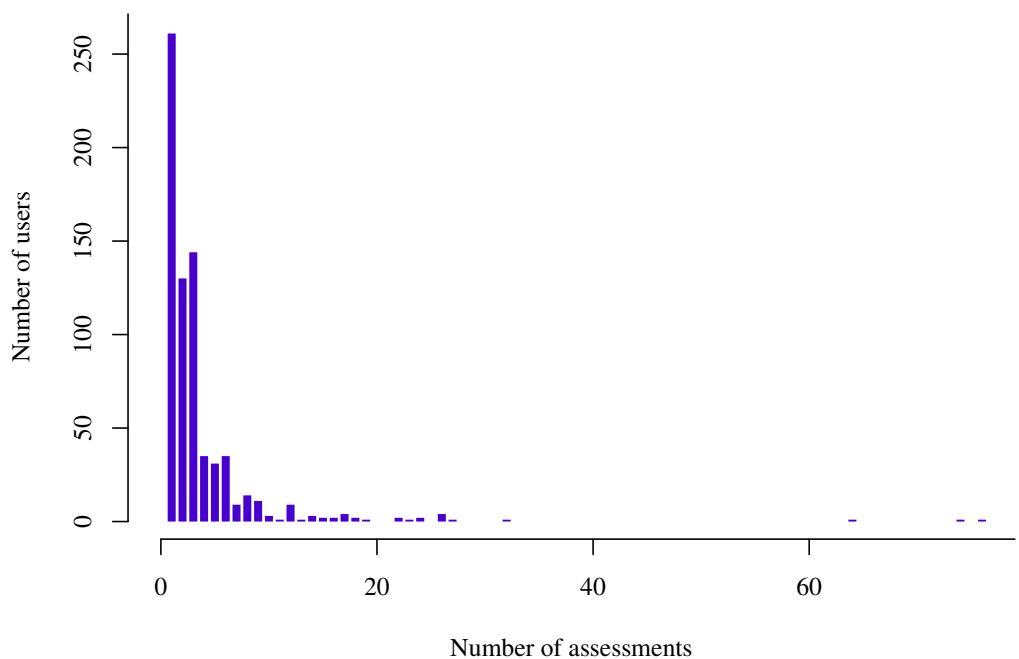
# A | APPENDIX

## A.1 EVALUATION OF MERCATO LINGUISTICO AND POKER PAROLE

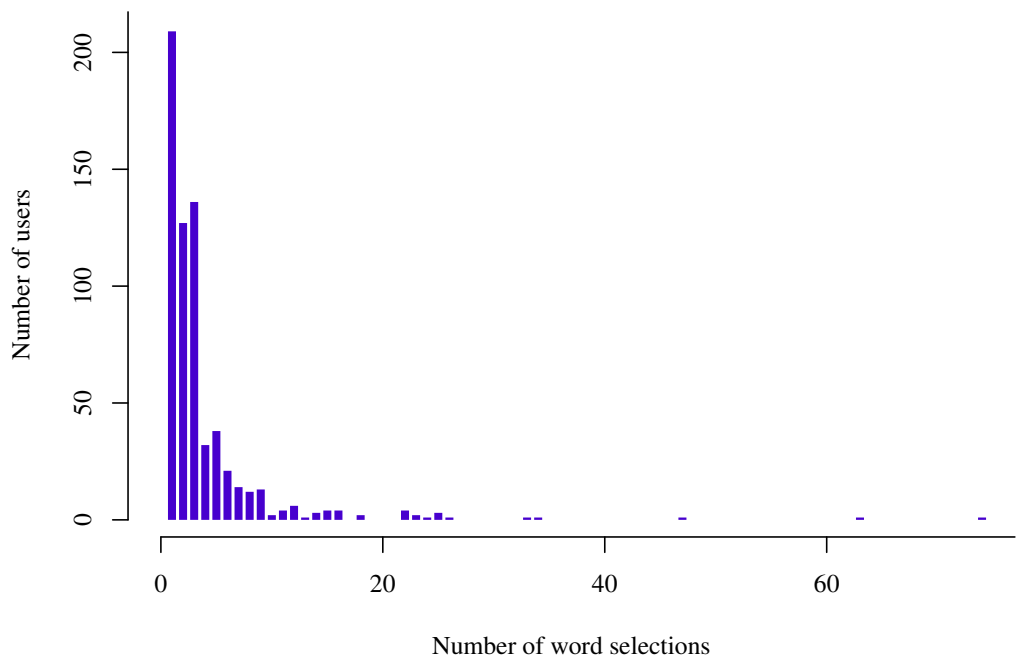
### A.1.1 Gathered Data per User



**Figure 35:** Distribution of the number of phrases added by single users on Mercato Linguistico and Poker Parole.

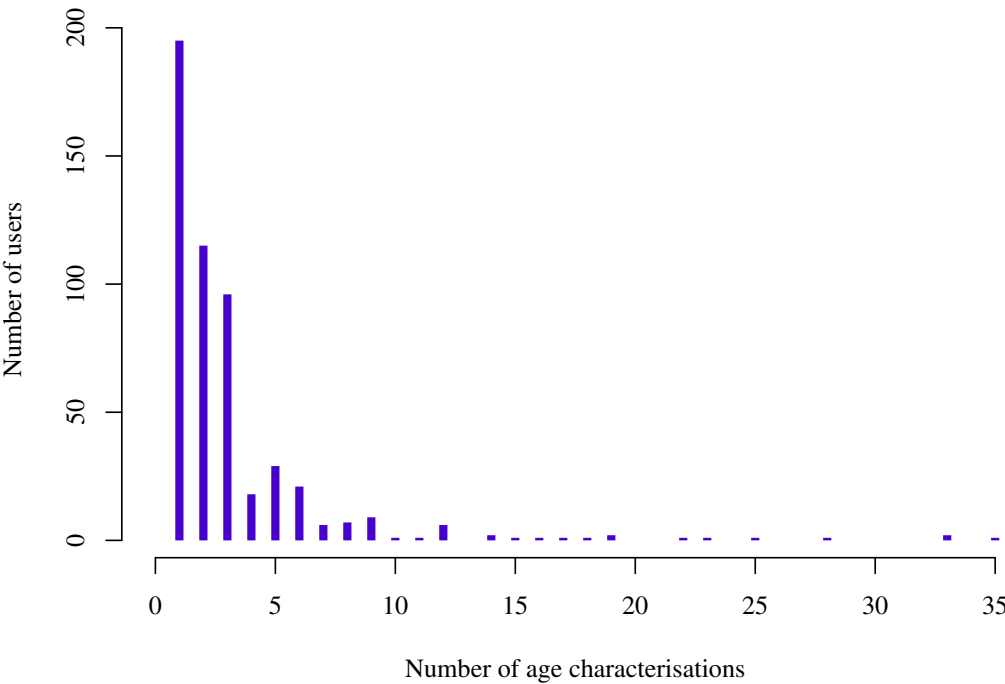


**Figure 36:** Distribution of the number of geographical assessments generated by single users on Mercato Linguistico.

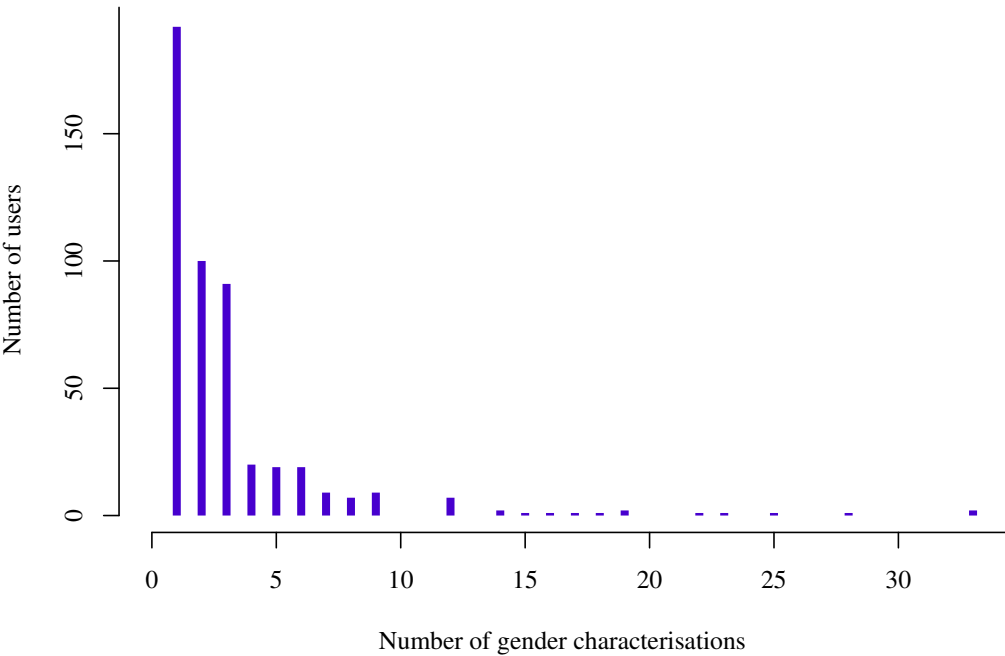


**Figure 37:** Distribution of the number of word selection characterisations generated by single users on Mercato Linguistico.

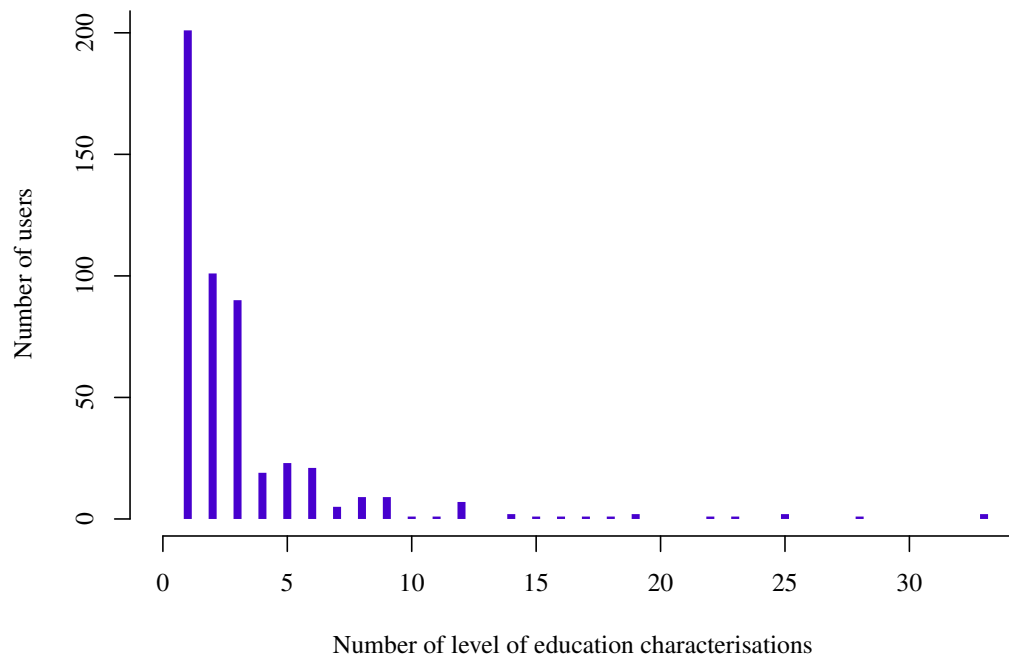




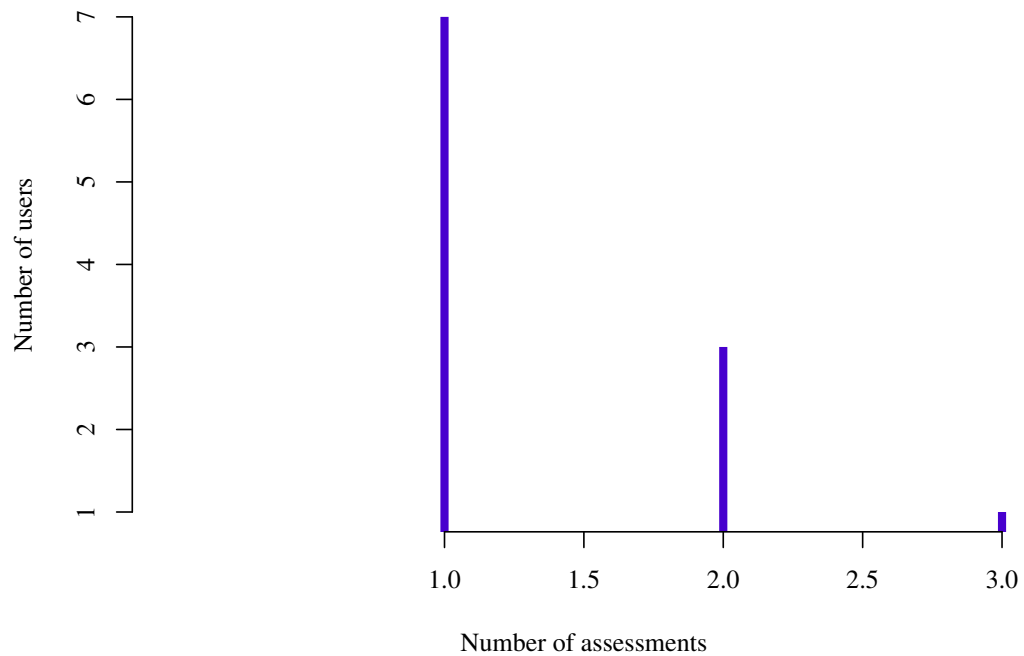
**Figure 38:** Distribution of the number of age characterisations generated by single users on Mercato Linguistico.



**Figure 39:** Distribution of the number of gender characterisations generated by single users on Mercato Linguistico.

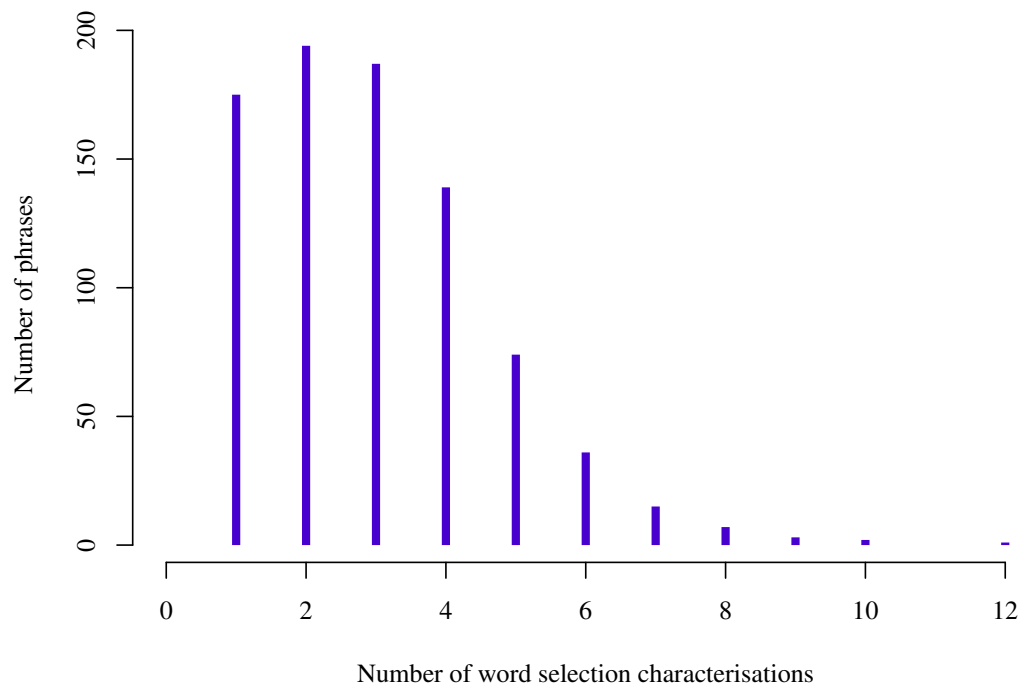


**Figure 40:** Distribution of the number of level of education characterisations generated by single users on Mercato Linguistico.

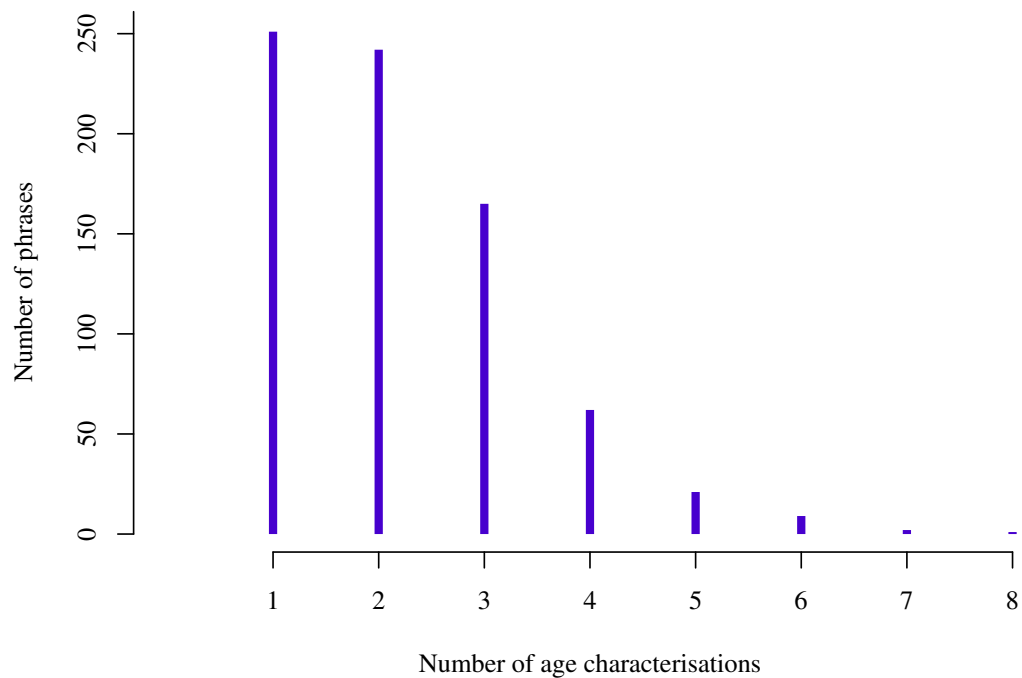


**Figure 41:** Distribution of the number of assessments generated by single users on Poker Parole.

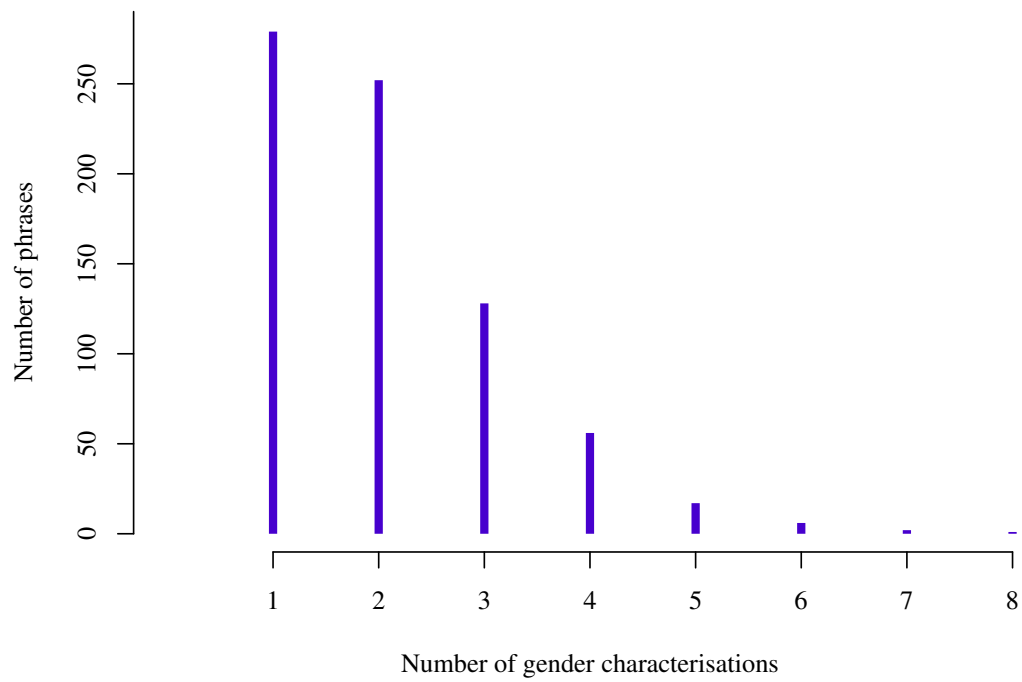
## A.1.2 Characterisations per Phrase



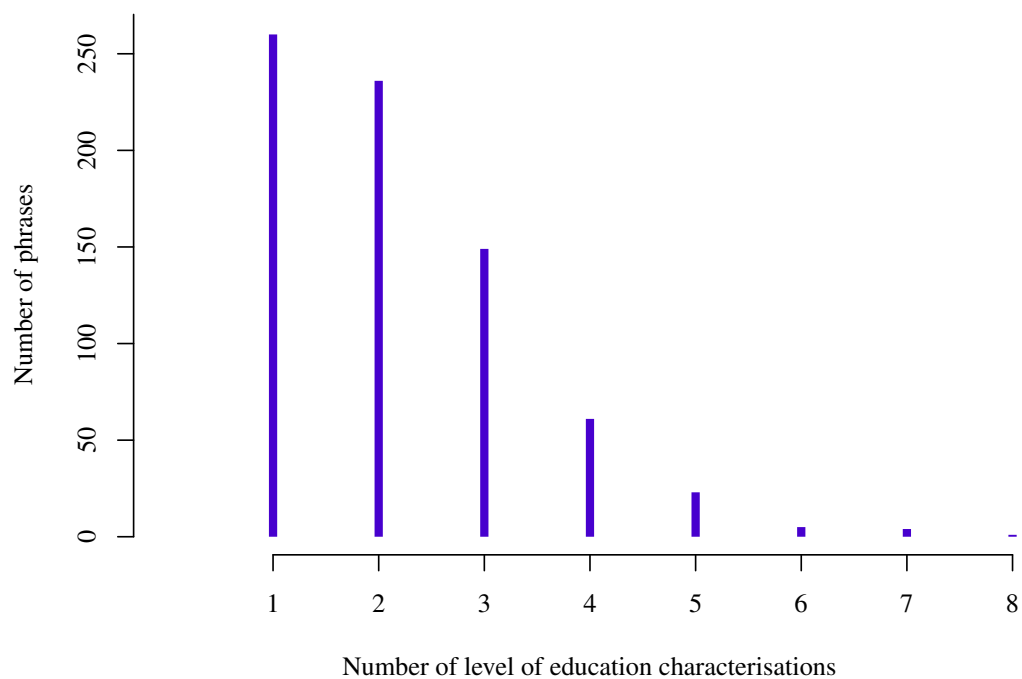
**Figure 42:** Distribution of the number of word selection characterisations per phrase on Mercato Linguistico.



**Figure 43:** Distribution of the number of age characterisations per phrase on Mercato Linguistico.

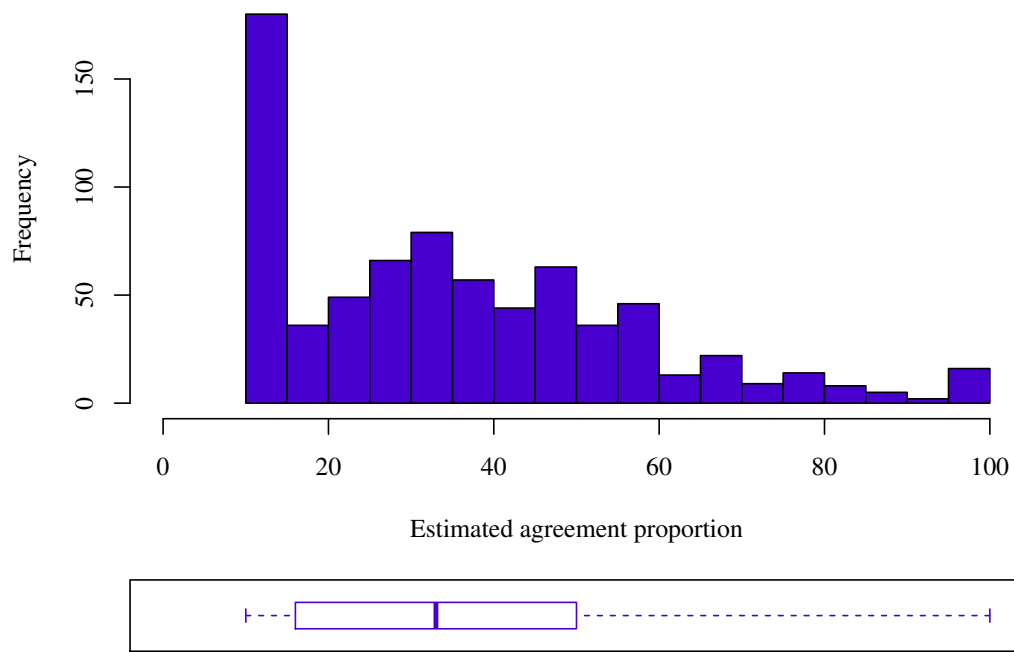


**Figure 44:** Distribution of the number of gender characterisations per phrase on Mercato Linguistico.



**Figure 45:** Distribution of the number of level of education characterisations per phrase on Mercato Linguistico.

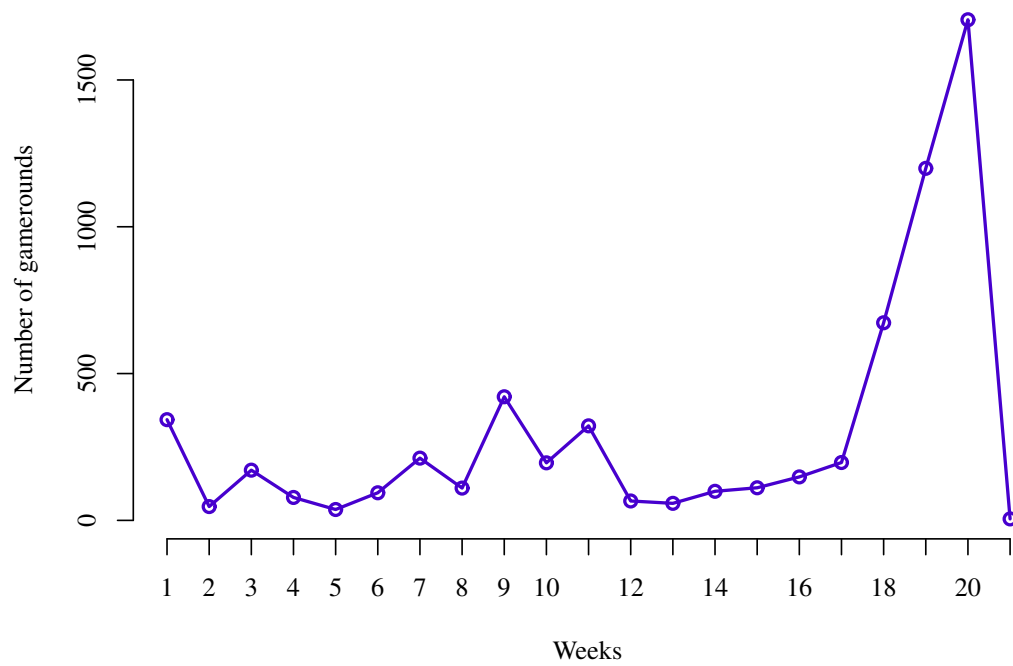
## A.1.3 The Assessment's Agreement Proportion



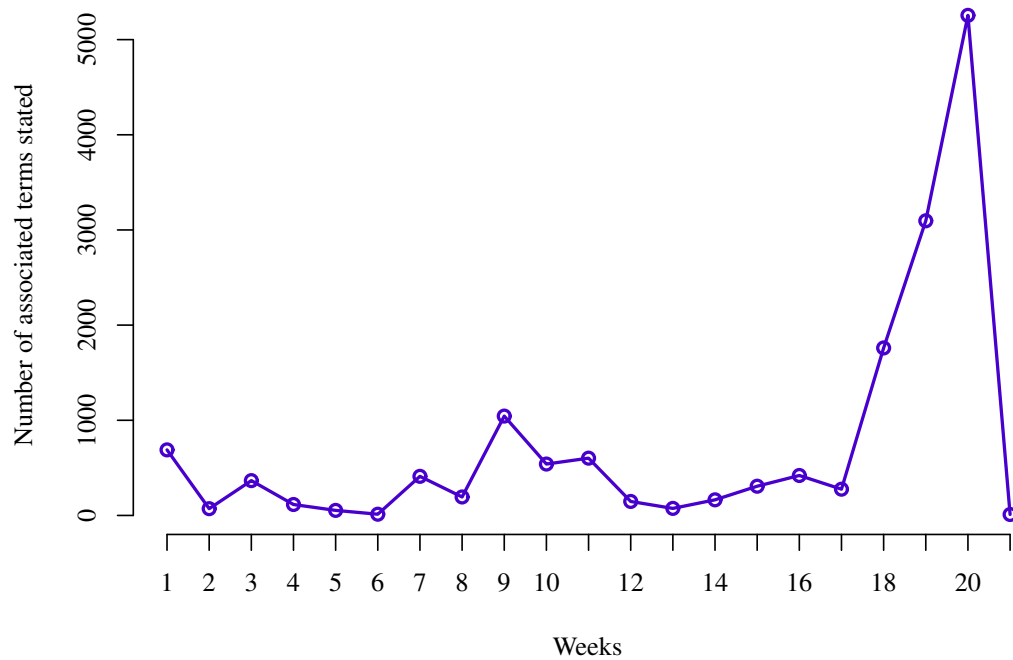
**Figure 46:** Distribution of the estimated agreement proportion of assessments with a difference between the estimated agreement proportion and the calculated agreement of at most 10 in Mercato Linguistico. The mean is 36 with a standard deviation of 22.

## A.2 EVALUATION OF TERMINA

### A.2.1 Gathered Data



**Figure 47:** Number of completed game rounds per week.

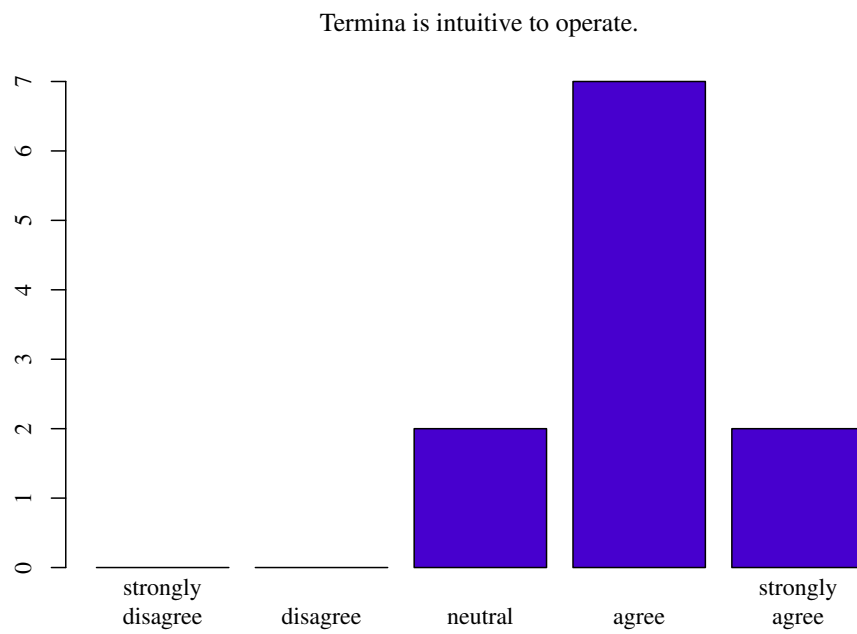
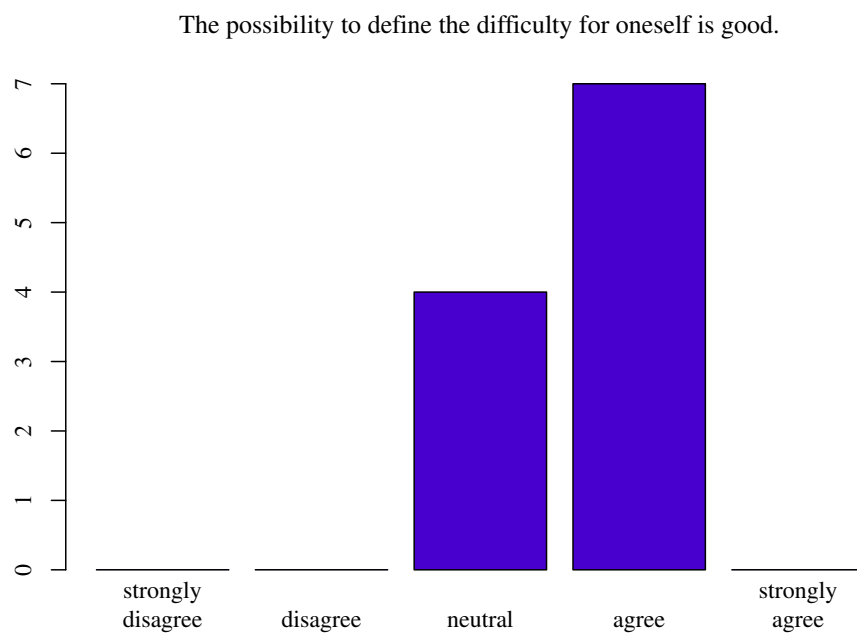


**Figure 48:** Number of stated associated terms per week.

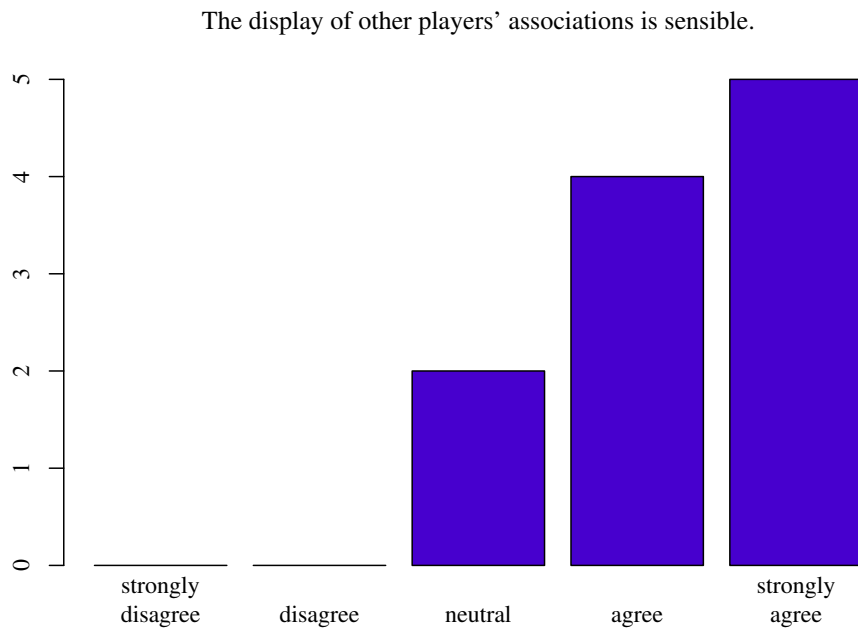
close	far	unclassified
Liste (20)	O (4)	-
list (13)	Assoziativ (3)	
Funktion (12)	Rekursion (2)	
SML (10)	zusammen (2)	
curried (6)		
Filter (6)		
höhere Ordnung (5)		
Map (5)		
Funktion höherer Ordnung (4)		
curry (3)		
fold (3)		
Funktionsanwendung (3)		
höhere funktion (3)		
unäre Funktion als Parameter (3)		
alle Listenelemente (2)		
anwenden (2)		
Element (2)		
foldl (2)		
foldr (2)		
iteration (2)		
Liste der Werte (2)		
Parameter (2)		
Polymorphie (2)		
unäre Funktion (2)		

**Table 7:** Associated terms, with their association strength in brackets, for the concept “map Funktion” (in English: “map function”) which have a minimum strength of two. This is the original version of the translated terms in [Table 5](#).

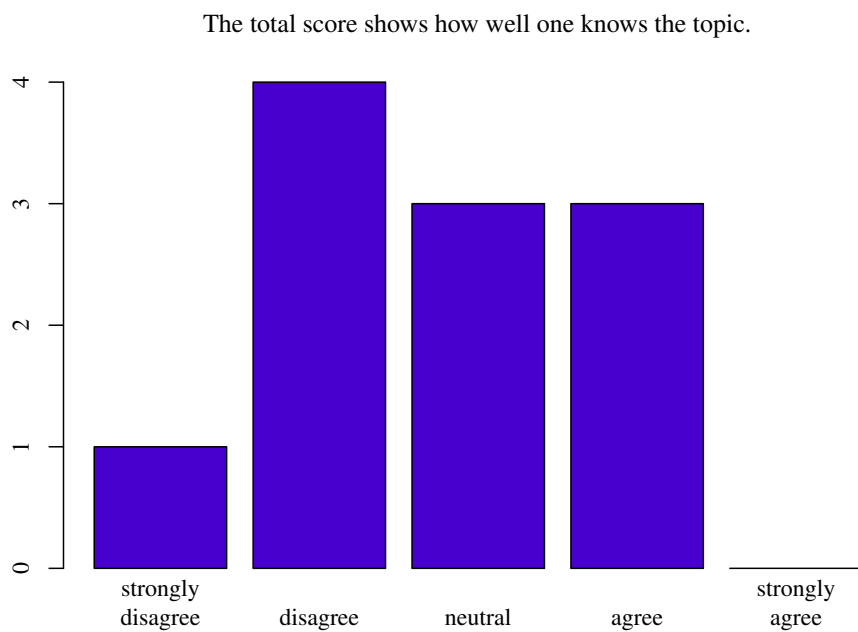
## A.2.2 User Survey

**Figure 49:** Answers to the survey question number 1.**Figure 50:** Answers to the survey question number 2.

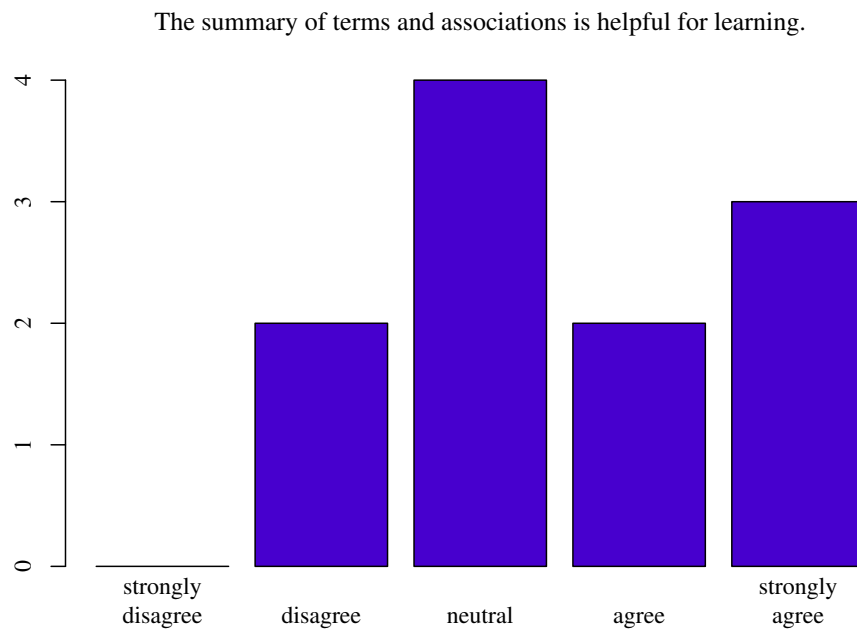




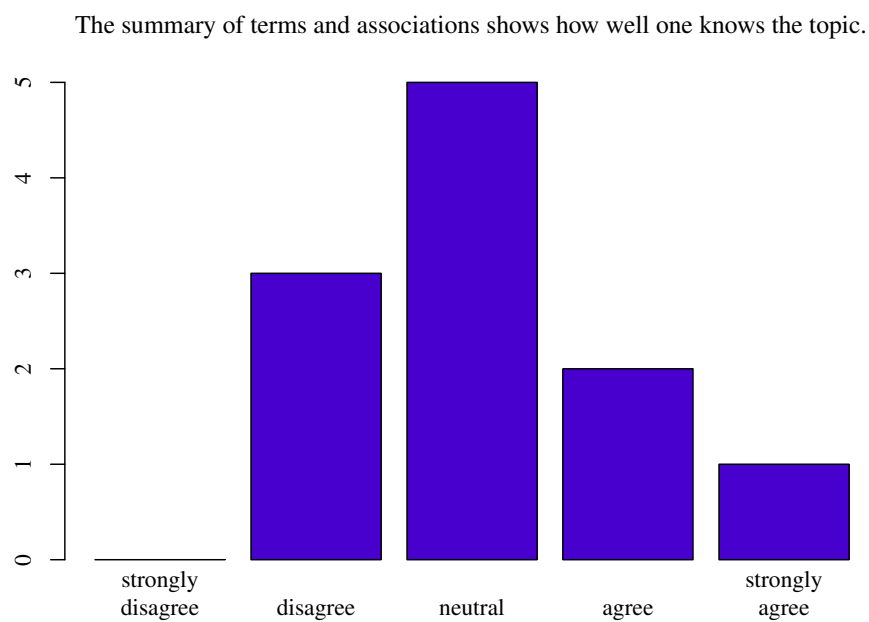
**Figure 51:** Answers to the survey question number 3.



**Figure 52:** Answers to the survey question number 4.



**Figure 53:** Answers to the survey question number 5.



**Figure 54:** Answers to the survey question number 6.

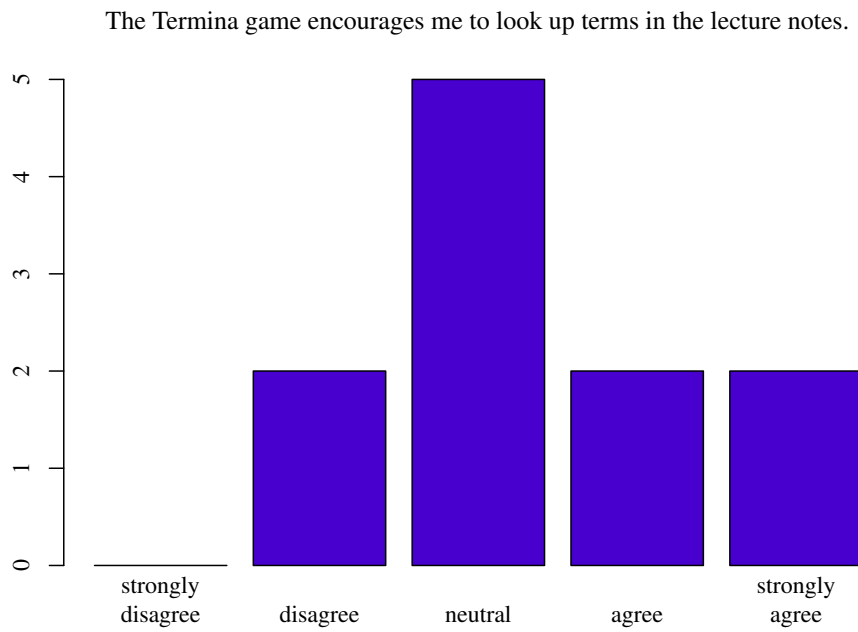


Figure 55: Answers to the survey question number 7.

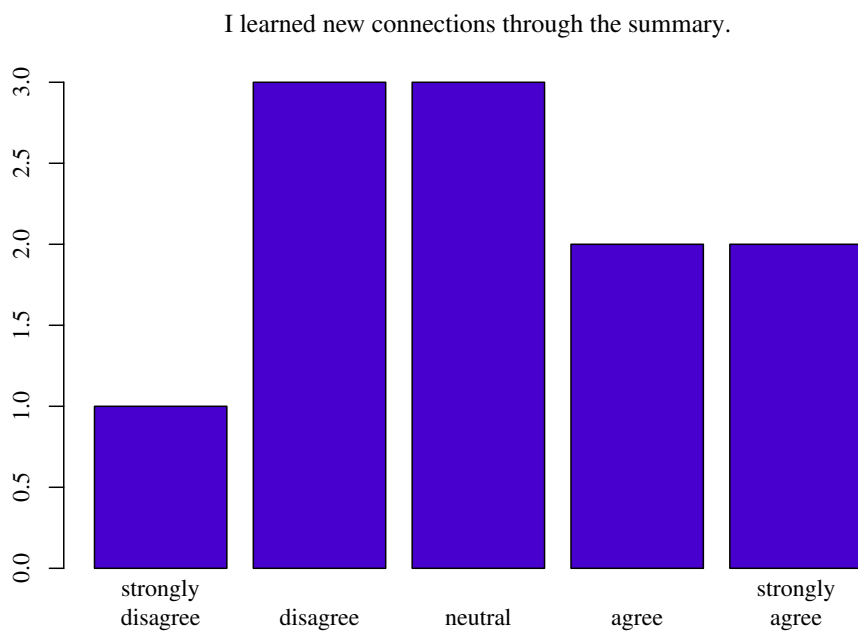


Figure 56: Answers to the survey question number 8.

Answers to the free form question “Hast du weitere Anmerkungen zum Termina Spiel?”:

- Die Assoziation “schwierig” sollte nicht vom Dozenten bestätigt werden...  
Wenn dann vielleicht noch “komplex” oder “kompliziert”
- Habe es nur für webinfo gespielt.  
eure änderungen zu vorher sind nicht alle gut:  
kennt ihr die Regel, dass umfragen möglichst auf einen 13zoll bildschirm passen sollen, damit der teilnehmer nicht scollen muss? diese regel habt ihr im SPiel gebrochen. ich muss jedes mal runterscrollen, wenn eine neue runde beginnt. jedesmal! das mindert den “spaß”. bitte ändert das. der 13Zoll bildschirm ist eine gänge größe, auf kleinere bildschirme müsst ihr keine Rücksicht mehr nehmen ;) )  
an sich find ich aber die visualisierung mit grün rot... sehr gut! aber die Begriffe überdecken sich mit dem Startbegriff, das müsst ihr auch noch ändern, die kann man dann nicht mehr lesen.  
außerdem habt ihr die begriffe geändert. ihr fragt nicht nach CSS sondern nach CSS merkmalen... weiß noch nicht, ob ich das gut finde... da ist es schwer zu erahnen, was ihr jetzt eigtl wissen wollt...  
ein lerneffekt ist bei mir jetzt noch nicht eingetreten, weil sämtliche Begriffe noch nicht bestätigt wurden. habe euch das schonmal als kritik iwo geschrieben. wenn 80% meiner begriffe orange sind, dann lern ich nichts, könnt ihr nicht sinnvolle begriffe schon vorher selber eingeben und bestätigen? damit die datenbank nicht von null aufgebaut wird?
- Ich finde die Idee ein Lernspiel für eine Vorlesung zu machen sehr gut. Allerdings reicht es nicht nur Fachbegriffe miteinander in Verbindung zu bringen... (Man braucht gar nicht zu wissen was sie bedeuten, sondern nur das sie irgendetwas miteinander zu tun haben). Allerdings wäre ein gutes Lernspiel auch ein sehr großer Arbeitsaufwand. Evtl. Könnte man Termina bzw. Den Lernerfolg ein bisschen verbessern, indem man Definitionen von Fachbegriffen an bestimmten Stellen anzeigt. Oder tatsächlich verständnisrelevante Multiple Choice Fragen einbaut.  
Ansonsten Vielen Dank für die Entwicklung des Spiels!
- Vielen Dank für das lustige Spiel!

## BIBLIOGRAPHY

- [1] Michael Abramowicz. “Information Markets, Administrative Decision-making, and Predictive Cost-Benefit Analysis”. In: *The University of Chicago Law Review* 71.3 (2004), pages 933–1020 (cited on page 15).
- [2] Clarac C Abt. *Serious Games*. New York: Viking, 1970 (cited on page 19).
- [3] Eytan Adar. “Why I Hate Mechanical Turk Research (and Workshops)”. In: *Proceedings of the 29th ACM Conference on Human Factors in Computing Systems (CHI)*. 2011 (cited on page 11).
- [4] Ahmet Aker, Mahmoud El-haj, M-dyaa Albakour and Udo Kruschwitz. “Assessing Crowdsourcing Quality through Objective Tasks”. In: *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC)*. 2012 (cited on page 11).
- [5] J. Albors, J.C. Ramos and J.L. Hervás. “New Learning Network Paradigms: Communities of Objectives, Crowdsourcing, Wikis and Open Source”. In: *International Journal of Information Management* 28.3 (2008), pages 194–202 (cited on page 10).
- [6] Omar Alonso, Daniel E Rose and Benjamin Stewart. “Crowdsourcing for Relevance Evaluation”. In: *ACM SIGIR Forum* 42.2 (2008), pages 9–15 (cited on page 10).
- [7] Kenneth J Arrow, Robert Forsythe, Michael Gorham, Robert Hahn, Robin Hanson, John O Ledyard, Saul Levmore, Robert Litan, Paul Milgrom, Forrest D Nelson, George R Neumann, Marco Ottaviani, Thomas C Schelling, Robert J Shiller, Vernon L Smith, Erik Snowberg, Cass R Sunstein, Paul C Tetlock, Philip E Tetlock, Hal R Varian, Justin Wolfers and Eric Zitzewitz. “The Promise of Prediction Markets”. In: *Science* 320 (2008), pages 877–878 (cited on page 15).
- [8] Jill Avery. “Leveraging Crowdsourced Peer-To-Peer Assessments to Enhance the Case Method of Learning”. In: *Journal for Advancement of Marketing Education* 22.1 (2014) (cited on page 3).
- [9] Marcia J. Bates. “What is Browsing - Really? A Model Drawing from Behavioural Science Research”. In: *Information Research* 12.4 (2007) (cited on page 49).
- [10] Tara S Behrend, David J Sharek, Adam W Meade and Eric N Wiebe. “The Viability of Crowdsourcing for Survey Research”. In: *Behavior Research Methods* 43.3 (2011), pages 800–813 (cited on page 3).
- [11] Francesco Bellotti, Bill Kapralos, Kiju Lee, Pablo Moreno-Ger and Riccardo Berta. “Assessment in and of Serious Games: An Overview”. In: *Advances in Human-Computer Interaction* 2013 (2013) (cited on page 101).

- [12] Joyce E Berg and Thomas A Rietz. “Prediction Markets as Decision Support Systems”. In: *Information Systems Frontiers* 5.1 (2003), pages 79–93 (cited on page 15).
- [13] Joyce Berg, Forrest Nelson and Thomas Rietz. *Accuracy and Forecast Standard Error of Prediction Markets*. Working Draft. University of Iowa, 2003 (cited on page 15).
- [14] Joyce Berg, Forrest Nelson and Thomas Rietz. “Prediction Market Accuracy in the Long Run”. In: *International Journal of Forecasting* 24.2 (2008), pages 283–298 (cited on page 15).
- [15] Michael S Bernstein, Joel Brandt, Robert C Miller and David R Karger. “Crowds in Two Seconds: Enabling Realtime Crowd-Powered Interfaces”. In: *Proceedings of the 24th Symposium on User Interface Software and Technology (UIST)*. 2011 (cited on page 11).
- [16] Michael S Bernstein, Greg Little, Robert C Miller, Björn Hartmann, Mark S Ackerman, David R Karger, David Crowell and Katrina Panovich. “Soylent: A Word Processor with a Crowd Inside”. In: *Proceedings of the 24th Symposium on User Interface Software and Technology (UIST)*. 2010 (cited on page 17).
- [17] Pier Marco Bertinetto. “La sillabazione dei nessi /sC/ in italiano: Un’eccezione alla tendenza ‘universale’?” In: *Atti de XXXI Congresso Internazionale di Studi della Società di Linguistica Italiana, "Fonologia e morfologia dell’italiano e dei suoi dialetti"*. 1999, pages 12–28 (cited on pages 27, 86).
- [18] Sue Blackman. “Serious Games... and less!” In: *ACM SIGGRAPH Computer Graphics* 39.1 (2005), pages 12–16 (cited on page 19).
- [19] David Bodoff and Eran Vaknin. “Don’t Look Here: Off-Limits Words Bias Play in the ESP Game”. In: *Proceedings of the Pacific Asia Conference on Information Systems (PACIS)*. 2013 (cited on page 13).
- [20] Eric Bonabeau. “Decisions 2.0: The Power of Collective Intelligence”. In: *MIT Sloan Management Review* 50.2 (2009), pages 45–52 (cited on page 8).
- [21] D. C. Brabham. “Crowdsourcing as a Model for Problem Solving: An Introduction and Cases”. In: *Convergence: The International Journal of Research into New Media Technologies* 14.1 (2008), pages 75–90 (cited on pages 10, 11).
- [22] Dominique Brossard, Bruce Lewenstein and Rick Bonney. “Scientific Knowledge and Attitude Change: The Impact of a Citizen Science Project”. In: *Science Education* 27.9 (2005), pages 1099–1121 (cited on page 19).
- [23] François Bry. “Human Computation-Enabled Network Analysis for a Systemic Credit Risk Rating”. In: *Handbook of Human Computation*. Edited by Pietro Michelucci. Springer, 2013 (cited on page 49).

- [24] François Bry, Fabian Kneissl, Thomas Krefeld, Stephan Lücke and Christoph Wieser. “Crowdsourcing for a Geographical and Social Mapping of Italian Dialects”. In: *Proceedings of the 2nd International Workshop on Social Media for Crowdsourcing and Human Computation at ACM Web Science (SoHuman)*. 2013 (cited on pages [4](#), [7](#), [25](#), [31](#), [55](#), [75](#)).
- [25] François Bry, Fabian Kneissl and Christoph Wieser. “Field Research for Humanities with Social Media: Crowdsourcing and Algorithmic Data Analysis”. In: *Proceedings of the 4th Workshop Digital Social Networks*. Berlin, 2011 (cited on pages [3](#), [7](#), [14](#), [73](#)).
- [26] François Bry and Christoph Wieser. “Squaring and Scripting the ESP Game: Trimming a GWAP to Deep Semantics”. In: *Proceedings of the International Conference on Serious Games Development and Applications*. 2012 (cited on page [14](#)).
- [27] Romy Buchschmid. “Development and Evaluation of a User Interface for an E-Learning Game”. Bachelor Thesis. Ludwig-Maximilians-Universität München, 2012. (Supervised by François Bry and Fabian Kneissl) (cited on page [105](#)).
- [28] J Burke, D Estrin, M Hansen, A Parker, N Ramanathan, S Reddy and M B Srivastava. “Participatory Sensing”. In: *Proceedings of the First Workshop on World-Sensor-Web: Mobile Device Centric Sensory Networks and Applications (WSW)*. 2006 (cited on page [12](#)).
- [29] Alberto J. Cañas and Joseph D. Novak. “Facilitating the Adoption of Concept Mapping Using CmapTools to Enhance Meaningful Learning”. In: *Knowledge Cartography, Software Tools and Mapping Techniques*. Edited by A. Okada, S. J. Buckinham Shum and T. Sherborne. Springer, 2008 (cited on page [22](#)).
- [30] Jon Chamberlain, Karen Fort, Udo Kruschwitz, Mathieu Lafourcade and Massimo Poesio. “Using Games to Create Language Resources: Successes and Limitations of the Approach”. In: *The People’s Web Meets NLP*. Edited by Iryna Gurevych and Jungi Kim. Springer, 2013, pages 3–44 (cited on page [17](#)).
- [31] Peter Pin-Shan Chen. “The Entity-Relationship Unified View of Data”. In: *ACM Transactions on Database Systems* 1 (1976), pages 9–36 (cited on page [122](#)).
- [32] Sandra L Chen and David R Michael. *Serious Games: Games That Educate, Train, and Inform*. Cengage Learning EMEA, 2005 (cited on page [19](#)).
- [33] Yen Yu Chen, Qingqing Gan and Torsten Suel. “I/O-Efficient Techniques for Computing PageRank”. In: *Proceedings of the 11th International Conference on Information and Knowledge Management (CIKM)*. McLean, Virginia, USA, 2002, pages 549–557 (cited on page [15](#)).

- [34] Yiling Chen, Chao-hsien Chu, Tracy Mullen and David M Pennock. “Information Markets vs. Opinion Pools: An Empirical Comparison”. In: *Proceedings of the 6th ACM Conference on Electronic Commerce (EC)*. 2005, pages 58–67 (cited on page 15).
- [35] Adrian David Cheok, Xubo Yang, Zhou Zhi Ying, Mark Billingham and Hirokazu Kato. “Touch-Space: Mixed Reality Game Space Based on Ubiquitous, Tangible, and Social Computing”. In: *Personal and Ubiquitous Computing* 6 (2002), pages 430–442 (cited on page 9).
- [36] Coye Cheshire and Judd Antin. “The Social Psychological Effects of Feedback on the Production of Internet Information Pools”. In: *Computer-Mediated Communication* 13.3 (2008), pages 705–727 (cited on page 70).
- [37] Christine Chin and David E. Brown. “Learning in Science: A Comparison of Deep and Surface Approaches”. In: *Journal of Research in Science Teaching* 37.2 (2000), pages 109–138 (cited on page 99).
- [38] Timothy Chklovski. “1001 Paraphrases: Incenting Responsible Contributions in Collecting Paraphrases from Volunteers”. In: *Proceedings of AAAI Spring Symposium: Knowledge Collection from Volunteer Contributors (KVC)*. 2005 (cited on page 17).
- [39] Alan Ivan Chorney. “Taking The Game Out Of Gamification”. In: *Dalhousie Journal of Interdisciplinary Management* 8.1 (2012) (cited on page 20).
- [40] Gianrenzo P. Clivio, Marcel Danesi and Sara Maida-Nicol. *An Introduction to Italian Dialectology*. Munich, Germany: LINCOM Europa, 2011 (cited on page 27).
- [41] Laura Commare. “Social Tagging als Methode zur Optimierung kunsthistorischer Bilddatenbanken – Eine Empirische Analyse des ARTigo-Projekts”. In: *Kunstgeschichte. Open Peer Reviewed Journal* (2011) (cited on page 25).
- [42] Thomas M. Connolly, Elizabeth a. Boyle, Ewan MacArthur, Thomas Hailey and James M. Boyle. “A Systematic Literature Review of Empirical Evidence on Computer Games and Serious Games”. In: *Computers & Education* 59.2 (2012), pages 661–686 (cited on page 19).
- [43] Seth Cooper, Adrien Treuille, Janos Barbero, Andrew Leaver Fay, Kathleen Tuite, Firas Khatib, Alex Cho Snyder, Michael Beenen, David Salesin, David Baker and Zoran Popović. “The Challenge of Designing Scientific Discovery Games”. In: *Proceedings of the 5th International Conference on the Foundation of Digital Games (FDG)*. 2010, pages 40–47 (cited on page 13).
- [44] Clara Pereira Coutinho. “Individual versus Collaborative Computer-Supported Concept Mapping: A Study with Adult Learners”. In: *Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education*. Vancouver, 2009, pages 1173–1180 (cited on page 22).



- [45] Bo Cowgill, Justin Wolfers and Eric Zitzewitz. “Using Prediction Markets to Track Information Flows: Evidence from Google”. In: *Proceedings of the 1st International ICST Conference*. Boston, MA, USA, 2009 (cited on page 15).
- [46] John Daniel. “Making Sense of MOOCs: Musings in a Maze of Myth, Paradox and Possibility”. In: *Journal of Interactive Media in Education* 3 (2012) (cited on page 21).
- [47] Charles Darrow and Elizabeth Magie. *Monopoly [Board Game]*. 1933 (cited on page 70).
- [48] Rachel Davies and Liz Sedley. *Agile Coaching*. Pragmatic Bookshelf, 2009 (cited on page 117).
- [49] Kathryn A. Davis. “Qualitative Theory and Methods in Applied Linguistics Research”. In: *TESOL Quarterly* 29.3 (1995), pages 427–453 (cited on page 27).
- [50] Tullio De Mauro. *Storia linguistica dell’Italia unita*. Laterza, 1963 (cited on page 27).
- [51] Sebastian Deterding, Dan Dixon, Rilla Khaled and Lennart Nacke. “From Game Design Elements to Gamefulness: Defining “Gamification””. In: *Proceedings of the 15th International Academic MindTrek Conference*. 2011, pages 9–15 (cited on page 20).
- [52] Dominic DiPalantino and Milan Vojnovic. “Crowdsourcing and All-Pay Auctions”. In: *Proceedings of the 10th ACM Conference on Electronic Commerce (EC)*. Stanford, California, USA, 2009, pages 119–128 (cited on page 11).
- [53] Anhai Doan, Raghu Ramakrishnan and Alon Y. Halevy. “Crowdsourcing Systems on the World-Wide Web”. In: *Communications of the ACM* 54.4 (2011), pages 86–96 (cited on pages 10, 11).
- [54] D. C. Dryer, C. Eisbach and W. S. Ark. “At What Cost Pervasive? A Social Computing View of Mobile Computing Systems”. In: *IBM Systems Journal* 38.4 (1999), pages 652–676 (cited on pages 8, 9).
- [55] Erik Duval. “Attention Please! Learning Analytics for Visualization and Recommendation”. In: *Proceedings of the 1st International Conference on Learning Analytics and Knowledge (LAK)*. 2011 (cited on page 21).
- [56] Erik Duval and Katrien Verbert. “Learning Analytics”. In: *eleed* 8 (2012) (cited on page 21).
- [57] Jacob Eisenstein, Brendan O Connor, Noah A Smith and Eric P Xing. “A Latent Variable Model for Geographic Lexical Variation”. In: *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2010, pages 1277–1287 (cited on page 17).
- [58] Thomas Erickson. “Social Computing”. In: *The Encyclopedia of Human-Computer Interaction*. Edited by Mads Soegaard and Rikke Friis Dam. 2nd. The Interaction Design Foundation, 2013 (cited on page 8).

- [59] Robert Forsythe, Forrest Nelson, George R Neumann and Jack Wright. “Anatomy of an Experimental Political Stock Market”. In: *The American Economic Review* 82.5 (1992), pages 1142–1161 (cited on page 17).
- [60] Michael J Franklin, Donald Kossmann, Tim Kraska, Sukriti Ramesh and Reynold Xin. “CrowdDB: Answering Queries with Crowdsourcing”. In: *Proceedings of the ACM International Conference on Management of Data (SIGMOD)*. Athens, Greece, 2011, pages 61–72 (cited on page 10).
- [61] Nigel R Franks. “Army Ants: A Collective Intelligence”. In: *American Scientist* 77.2 (1989), pages 138–145 (cited on page 7).
- [62] Thomas M. J. Fruchterman and Edward M. Reingold. “Graph Drawing by Force-Directed Placement”. In: *Software-Practice and Experience* 21.11 (1991), pages 1129–1164 (cited on page 103).
- [63] D Randy Garrison. *E-Learning in the 21st Century. A Framework for Research and Practice*. Routledge, 2011 (cited on page 21).
- [64] Benjamin M Good and Andrew I Su. “Games with a Scientific Purpose”. In: *Genome Biology* 12 (2011) (cited on pages 3, 13).
- [65] Michael F. Goodchild and J. Alan Glennon. “Crowdsourcing Geographic Information for Disaster Response: A Research Frontier”. In: *International Journal of Digital Earth* 3.3 (2010), pages 231–241 (cited on page 10).
- [66] Thavamalar Govindasamy. “Successful Implementation of E-Learning: Pedagogical Considerations”. In: *The Internet and Higher Education* 4.3-4 (2001), pages 287–299 (cited on page 21).
- [67] James G. Greeno, Allan M. Colins and Lauren B. Resnick. “Cognition and Learning”. In: *Handbook of Educational Psychology*. Edited by D. C. Berliner and R. C. Calfee. New York: Macmillan, 1996, pages 15–46 (cited on page 99).
- [68] David Alan Grier. *When Computers Were Human*. Princeton University Press, 2005 (cited on page 12).
- [69] Fabian Groh. “Gamification: State of the Art Definition and Utilization”. In: *Proceedings of the 4th Seminar on Research Trends in Media Informatics*. 2012, pages 39–46 (cited on page 20).
- [70] Georg Groh, Michele Brocco and Yonata A Asikin. “Contribution Awareness and Fame in Open Innovation Networks”. In: *Proceedings of the International Symposium on Support for Open Innovation Processes*. 2009 (cited on page 45).
- [71] Adelaide Haas. “Male and female spoken language differences: Stereotypes and evidence.” In: *Psychological Bulletin* 86.3 (1979), pages 616–626 (cited on page 29).

- [72] Hicham Hage. “Harnessing Learner’s Collective Intelligence: A Web2.0 Approach to E-Learning”. In: *Intelligent Tutoring Systems*. Edited by Beverly P. Woolf, Esma Aimeur, Roger Nkambou and Susanne Lajoie. Lecture Notes in Computer Science. Springer, 2008, pages 438–447 (cited on page 21).
- [73] Robert W Hahn and Paul C Tetlock. “Using Information Markets to Improve Public Decision Making”. In: *Harvard Journal of Law and Public Policy* 29.1 (2005), pages 213–289 (cited on page 15).
- [74] Nils H. Hakansson, Avraham Beja and Jivendra Kale. “On the Feasibility of Automated Market Making by a Programmed Specialist”. In: *The Journal of Finance* 11.1 (1985) (cited on page 43).
- [75] Azma Abdul Hamid. “E-Learning: Is it the “e” or the Learning that Matters?” In: *The Internet and Higher Education* 4.3-4 (2001), pages 311–316 (cited on page 21).
- [76] Robin Hanson. “Idea Futures: Encouraging an Honest Consensus”. In: *Extropy* 3.2 (1992), pages 7–17 (cited on page 15).
- [77] Einar Haugen. “Dialect, Language, Nation”. In: *American Anthropologist* 68.4 (1966), pages 922–935 (cited on page 29).
- [78] Jutta Heckhausen and Heinz Heckhausen. *Motivation and Action*. Cambridge University Press, 2008 (cited on pages 45, 46, 70).
- [79] Alan R Hevner, Salvatore T March, Jinsoo Park and Sudha Ram. “Design Science in Information Systems Research”. In: *MIS Quarterly* 28.1 (2004), pages 75–105 (cited on page 3).
- [80] Birger Hjørland. “The Importance of Theories of Knowledge: Browsing as an Example”. In: *Journal of the American Society for Information Science and Technology* 62.3 (2011), pages 594–603 (cited on page 49).
- [81] Birger Hjørland. “Theoretical Clarity is not “Manicheanism”: A Reply to Marcia Bates”. In: *Journal of Information Science* 37.5 (2011), pages 546–550 (cited on page 49).
- [82] Barbora Hladká, Jiří Mírovský and Pavel Schlesinger. “Designing a Language Game for Collecting Coreference Annotation”. In: *Proceedings of the 3rd Linguistic Annotation Workshop (ACL-IJCNLP)*. 2009, pages 52–55 (cited on page 17).
- [83] Chien-Ju Ho and Kuan-Ta Chen. “On formal models for social verification”. In: *Proceedings of the ACM SIGKDD Workshop on Human Computation (HCOMP)*. 2009, pages 62–69 (cited on page 14).
- [84] John Joseph Horton and Lydia B. Chilton. “The Labor Economics of Paid Crowdsourcing”. In: *Proceedings of the 11th ACM Conference on Electronic Commerce (EC)*. New York, New York, USA, 2010, pages 209–218 (cited on page 11).

- [85] Gary Hsieh and Scott Counts. “mimir: A Market-Based Real-Time Question and Answer Service”. In: *Proceedings of the 27th ACM Conference on Human Factors in Computing Systems (CHI)*. Boston, MA, USA, 2009, pages 769–778 (cited on page 18).
- [86] Gary Hsieh, Robert E Kraut and Scott E Hudson. “Why Pay?: Exploring How Financial Incentives are Used for Question & Answer”. In: *Proceedings of the 28th ACM Conference on Human Factors in Computing Systems (CHI)*. 2010 (cited on page 11).
- [87] Gary Hsieh, Robert Kraut, Scott E Hudson and Roberto Weber. “Can Markets Help?: Applying Market Mechanisms to Improve Synchronous Communication”. In: *Proceedings of the Conference on Computer Supported Cooperative Work (CSCW)*. San Diego, California, USA, 2008, pages 535–544 (cited on page 18).
- [88] Pei-yun Hsueh, Prem Melville and Vikas Sindhwani. “Data Quality from Crowdsourcing: A Study of Annotation Selection Criteria”. In: *Proceedings of the NAACL HLT Workshop on Active Learning for Natural Language Processing*. Boulder, Colorado, USA, 2009, pages 27–35 (cited on page 11).
- [89] Shih-Wen Huang and Wai-Tat Fu. “Systematic Analysis of Output Agreement Games”. In: *Proceedings of the AAAI Workshop on Human Computation (HCOMP)*. 2012, pages 2–8 (cited on page 14).
- [90] Paul Hudak, John Hughes, Simon Peyton Jones and Philip Wadler. “A History of Haskell: Being Lazy With Class”. In: *Proceedings of the 3rd ACM SIGPLAN Conference on History of Programming Languages (HOPL III)*. 2007 (cited on pages 107, 108).
- [91] John Hull. “Options, Futures and Other Derivatives”. In: *Management*. Boston, MA, USA: Pearson, 2012 (cited on page 49).
- [92] Kai Huotari and Juho Hamari. “Defining Gamification - A Service Marketing Perspective”. In: *Proceedings of the 16th International Academic MindTrek Conference*. 2012 (cited on page 20).
- [93] Panagiotis G. Ipeirotis, Foster Provost and Jing Wang. “Quality management on Amazon Mechanical Turk”. In: *Proceedings of the ACM SIGKDD Workshop on Human Computation (HCOMP)*. New York, New York, USA, 2010 (cited on page 11).
- [94] Shaili Jain and David Parkes. “A Game-Theoretic Analysis of Games with a Purpose”. In: *Proceedings of the 4th International Workshop on Internet and Network Economics (WINE)*. 2008, pages 342–350 (cited on page 14).
- [95] Shaili Jain and David C Parkes. “The Role of Game Theory in Human Computation Systems”. In: *Proceedings of the ACM SIGKDD Workshop on Human Computation (HCOMP)*. 2009, pages 58–61 (cited on page 14).

- [96] David Jamae and Peter Johnson. *JBoss in Action*. Manning, 2009 (cited on page [119](#)).
- [97] Seven L. Jones and Jeffry M. Netter. “Efficient Capital Markets”. In: *The Concise Encyclopedia of Economic*. Edited by David R. Henderson. Liberty Fund, Inc., 2008 (cited on page [15](#)).
- [98] Uwe Jürgens. “Die Evolution der Sprache: Vom Affenlaut zum Menschenwort”. In: *Biologie in unserer Zeit* 36.6 (2006), pages 362–368 (cited on page [28](#)).
- [99] Epaminondas Kapetanios. “Quo Vadis Computer Science: From Turing to Personal Computer, Personal Content and Collective Intelligence”. In: *Data & Knowledge Engineering* 67.2 (2008), pages 286–292 (cited on page [8](#)).
- [100] Andreas M. Kaplan and Michael Haenlein. “Users of the World, Unite! The Challenges and Opportunities of Social Media”. In: *Business Horizons* 53.1 (2010), pages 59–68 (cited on page [10](#)).
- [101] Robert A. Jr. Karasek. “Job Demands, Job Decision Latitude, and Mental Strain: Implications for Job Redesign”. In: *Administrative Science Quarterly* 24.2 (2014), pages 285–308 (cited on page [95](#)).
- [102] John Maynard Keynes. *The General Theory of Employment, Interest, and Money*. Macmillan Cambridge University Press, 1936 (cited on pages [64](#), [65](#)).
- [103] Irwin King, Jiexing Li and Kam Tong Chan. “A Brief Survey of Computational Approaches in Social Computing”. In: *Proceedings of the International Joint Conference on Neural Networks*. Atlanta, Georgia, USA, 2009, pages 1625–1632 (cited on page [9](#)).
- [104] Aniket Kittur, Ed H Chi and Bongwon Suh. “Crowdsourcing User Studies With Mechanical Turk”. In: *Proceedings of the 26th ACM Conference on Human Factors in Computing Systems (CHI)*. Florence, Italy, 2008, pages 453–456 (cited on pages [10](#), [11](#)).
- [105] Aniket Kittur, Boris Smus, Susheel Khamkar and Robert E Kraut. “CrowdForge: Crowdsourcing Complex Work”. In: *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology (UIST)*. Santa Barbara, CA, USA, 2011, pages 43–52 (cited on page [11](#)).
- [106] Eric Klopfer, Scot Osterweil and Katie Salen. *Moving Learning Games Forward*. The Education Arcade. Massachusetts Institute of Technology, 2009 (cited on page [21](#)).
- [107] Fabian Kneissl and François Bry. *Borsa Parole - A Market for Linguistic Speculation*. Research Report. Institute for Informatics, Ludwig-Maximilian University of Munich, 2012 (cited on pages [7](#), [25](#), [55](#)).



- [108] Fabian Kneissl and François Bry. “Fostering Concept Maps Awareness as a Means to Learning”. In: *Proceedings of the Third International Conference on Social Computing and Its Applications (SCA)*. Karlsruhe, Germany, 2013 (cited on pages 7, 91).
- [109] Fabian Kneissl and François Bry. “Metropolitalia: A Crowdsourcing Platform for Linguistic Field Research”. In: *Proceedings of the IADIS International Conference WWW/Internet*. 2012 (cited on pages 7, 25, 55, 75).
- [110] Fabian Kneissl and François Bry. “Speculation and Bluff for Linguistic Field Research - Social Software and Media for Data Gathering”. In: *Proceedings of 5. Workshop Digitale Soziale Netze im Rahmen der 42. Jahrestagung der Gesellschaft für Informatik (GI)*. 2012 (cited on pages 7, 25, 31, 55).
- [111] Fabian Kneissl, Thomas Krefeld and Christoph Wieser. *A Crowdsourcing Platform for Italian Linguistic Field Research*. Research Report. Institute for Informatics, Ludwig-Maximilians-Universität München, 2013 (cited on pages 7, 25, 31, 55, 75).
- [112] Hubertus Kohle. “Social Tagging in the Humanities and the Nature of the Internet Workflow”. In: *Kunstgeschichte. Open Peer Reviewed Journal* (2011) (cited on pages 14, 25, 73).
- [113] Hubertus Kohle, François Bry, Thomas Krefeld, Christian Riepl and Klaus Schulz. *Abschlussbericht zum Projekt Entwicklung sozialer Web-Plattformen zur Datengewinnung in den Geisteswissenschaften (KO 1091/4-2)*. Technical report July 2011. 2013 (cited on page 73).
- [114] Thomas Krefeld. “Alter Standard - Neue Medien. Zur Erfassung von Restandardisierungsprozessen im Italienischen”. In: *Koineisierung und Standardisierung in der Romania*. Edited by Sarah Dessì Schmid, Jochen Hafner and Sabine Heinemann. Universitätsverlag Winter, 2011, pages 269–281 (cited on pages 28, 88).
- [115] Thomas Krefeld. “Italienische Varietätenlinguistik”. In: *Italienisch. Zeitschrift für italienische Sprache und Literatur* 63 (2010), pages 56–62 (cited on pages 27–29).
- [116] Katharina Krug. “Entwicklung und Evaluierung eines E-Learning Game With A Purpose”. Project thesis. Ludwig-Maximilians-Universität München, 2011. (Supervised by François Bry and Fabian Kneissl) (cited on page 105).
- [117] Ravi Kumar, Yury Lifshits and Andrew Tomkins. “Evolution of Two-Sided Markets”. In: *Proceedings of the third ACM International Conference on Web Search and Data Mining (WSDM)*. New York City, New York, USA, 2010 (cited on page 18).
- [118] Rachael Kwai Fun IP and Christian Wagner. “Weblogging: A Study of Social Computing and its Impact on Organizations”. In: *Decision Support Systems* 45.2 (2008), pages 242–250 (cited on page 8).

- [119] Mathieu Lafourcade. “Making People Play for Lexical Acquisition with the JeuxDeMots prototype”. In: *Proceedings of the 7th Symposium on Natural Language Processing (SNLP)*. 2007 (cited on page 17).
- [120] Walter S Lasecki, Young Chol Song, Henry Kautz and Jeffrey P Bigham. “Real-Time Crowd Labeling for Deployable Activity Recognition”. In: *Proceedings of the 16th ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW)*. 2013 (cited on page 11).
- [121] Terrance James Lavender. “Homeless: It’s No Game - Measuring the Effectiveness of a Persuasive Videogame”. Master’s Thesis. Simon Fraser University, 2008 (cited on page 19).
- [122] Edith Law, Burr Settles, Aaron Snook, Harshit Surana, Luis von Ahn and Tom Mitchell. “Human Computation for Attribute and Attribute Value Acquisition”. In: *Proceedings of the 1st Workshop on Fine-Grained Visual Categorization (FGVC)*. 2011 (cited on page 14).
- [123] Edith Law and Luis von Ahn. “Human Computation”. In: *Synthesis Lectures on Artificial Intelligence and Machine Learning*. Edited by Ronald J. Brachman, William W. Cohen and Thomas Dietterich. Morgan & Claypool Publishers, 2011, pages 1–121 (cited on pages 12–14).
- [124] Edith Law and Luis von Ahn. “Input-Agreement: A New Mechanism for Collecting Data Using Human Computation Games”. In: *Proceedings of the 27th ACM Conference on Human Factors in Computing Systems (CHI)*. 2009, pages 1197–1206 (cited on page 14).
- [125] Anne Lazaraton. “Qualitative Research in Applied Linguistics: A Progress Report”. In: *TESOL Quarterly* 29.3 (1995), pages 455–472 (cited on page 27).
- [126] Matthew Lease. “On Quality Control and Machine Learning in Crowdsourcing”. In: *Proceedings of the AAAI Workshop on Human Computation (HCOMP)*. 2011, pages 97–102 (cited on page 50).
- [127] Jan Marco Leimeister, Michael Huber, Ulrich Bretschneider and Helmut Krcmar. “Leveraging Crowdsourcing: Activation-Supporting Components for IT-Based Ideas Competition”. In: *Journal of Management Information Systems* 26.1 (2009), pages 197–224 (cited on page 10).
- [128] Anna Laura Lepschy and Guilio Lepschy. *The Italian Language Today*. 2nd. Routledge, 1988 (cited on pages 27, 29).
- [129] Stephan Leutenmayr, François Bry, Tom Schiebler and Felix Brodbeck. “Work in Progress: Do They Really Mean it? Assessing Decision Market Outcomes”. In: *Proceedings of the 4th Workshop Digital Social Networks*. 2011 (cited on page 16).
- [130] Stephan Leutenmayr, Fabian Kneissl and Sven Ziemer. “Gameful Markets for Collaboration and Learning”. In: *Proceedings of Disco, Workshop on Human Computation and Machine Learning in Games at HCOMP*. 2013 (cited on page 7).

- [131] Pierre Lévy. *Collective Intelligence: Mankind's Emerging World in Cyberspace*. Perseus Books, 1997. Chapter 277 (cited on pages 7, 8).
- [132] Greg Little, Lydia B Chilton, Max Goldman and Robert C Miller. "Exploring Iterative and Parallel Human Computation Processes". In: *Proceedings of the ACM SIGKDD Workshop on Human Computation (HCOMP)*. New York, New York, USA, 2010, pages 68–76 (cited on page 12).
- [133] Greg Little, Lydia B Chilton, Max Goldman and Robert C Miller. "TurKit: Tools for Iterative Tasks on Mechanical Turk". In: *Proceedings of the ACM SIGKDD Workshop on Human Computation (HCOMP)*. Paris, France, 2009, pages 29–30 (cited on page 11).
- [134] Yefeng Liu, Todorka Alexandrova and Tatsuo Nakajima. "Gamifying Intelligent Environments". In: *Proceedings of the International ACM Workshop on Ubiquitous Meta User Interfaces (Ubi-MUI)*. 2011, pages 7–12 (cited on page 20).
- [135] Yefeng Liu, Vili Lehdonvirta, Todorka Alexandrova and Tatsuo Nakajima. "Drawing on Mobile Crowds via Social Media". In: *Multimedia Systems* 18.1 (2012), pages 53–56 (cited on page 70).
- [136] E. B. Mallon, S. C. Pratt and N. R. Franks. "Individual and Collective Decision-Making During Nest Site Selection by the Ant *Leptothorax Albipennis*". In: *Behavioral Ecology and Sociobiology* 50.4 (2001), pages 352–359 (cited on page 7).
- [137] Thomas W Malone, Robert Laubacher and Chrysanthos Dellarocas. *Harnessing Crowds: Mapping the Genome of Collective Intelligence*. MIT Sloan Working Paper, number 4732-09. MIT Sloan School of Management, 2009 (cited on pages 7, 8).
- [138] David Maltz and Kate Ehrlich. "Pointing The Way: Active Collaborative Filtering". In: *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. 1995 (cited on page 71).
- [139] Winter Mason and Duncan J Watts. "Financial Incentives and the "Performance of Crowds"". In: *Proceedings of the ACM SIGKDD Workshop on Human Computation (HCOMP)*. 2009, pages 77–85 (cited on page 11).
- [140] Maja J Mataric. "Designing Emergent Behaviors: From Local Interactions to Collective Intelligence". In: *Proceedings of the Conference on Simulation of Adaptive Behavior*. Edited by J-A. Meyer, H. Roitblat and S. Wilson. MIT Press, 1993, pages 432–441 (cited on page 7).
- [141] Sebastian Matyas, Christian Matyas, Christoph Schlieder, Peter Kiefer, Hiroko Mitarai and Maiko Kamata. "Designing Location-Based Mobile Games with a Purpose: Collecting Geospatial Data with CityExplorer". In: *Proceedings of the International Conference on Advances in Computer Entertainment Technology (ACE)*. Yokohama, Japan, 2008, pages 244–247 (cited on page 13).



- [142] Jane McGonigal. *Reality is Broken*. The Penguin Press, 2011 (cited on page 19).
- [143] Luca Melchior. “«Frocio», «checca», «morosa» e... un problema lessicografico”. In: *Italienisch. Zeitschrift für italienische Sprache und Literatur* (2009), pages 67–88 (cited on pages 28, 29).
- [144] Paul R. Messinger, Eleni Stroulia, Kelly Lyons, Michael Bone, Run H. Niu, Kristen Smirnov and Stephen Perelgut. “Virtual Worlds - Past, Present, and Future: New Directions in Social Computing”. In: *Decision Support Systems* 47.3 (2009), pages 204–228 (cited on page 9).
- [145] Nolan Miller, Paul Resnick and Richard Zeckhauser. “Eliciting Informative Feedback: The Peer-Prediction Method”. In: *Management Science* 51.9 (2005), pages 1359–1373 (cited on page 14).
- [146] Patrick Minder and Abraham Bernstein. “CrowdLang: A Programming Language for the Systematic Exploration of Human Computation Systems”. In: *Social Informatics*. Edited by Karl Aberer, Andreas Flache, Wander Jager, Ling Liu, Jie Tang and Christophe Guéret. Springer, 2012, pages 124–137 (cited on pages 11, 17).
- [147] Patrick Minder, Sven Seuken, Abraham Bernstein and Mengia Zollinger. “CrowdManager - Combinatorial Allocation and Pricing of Crowdsourcing Tasks with Time Constraints”. In: *Proceedings of the Workshop on Social Computing and User Generated Content*. 2012 (cited on pages 11, 18).
- [148] A. M. Mioni and A. M. Arnuzzo-Lanszweert. “Sociolinguistics in Italy”. In: *International Journal of the Sociology of Language* 1979.21 (1979), pages 81–108 (cited on pages 27, 29).
- [149] Robert Munro, Steven Bethard, Victor Kuperman, Vicky Tzuyin Lai, Robin Melnick, Christopher Potts, Tyler Schnoebelen and Harry Tily. “Crowdsourcing and Language Studies: The new Generation of Linguistic Data”. In: *Proceedings of the Workshop on Creating Speech and Language Data with Amazon’s Mechanical Turk at NAACL-HLT*. 2010, pages 122–130 (cited on pages 3, 16).
- [150] Yiftach Nagar and Thomas W Malone. “Improving Predictions with Hybrid Markets”. In: *Proceedings of the AAAI Fall Symposium*. 2012, pages 30–36 (cited on page 17).
- [151] Joseph D. Novak and Alberto J. Cañas. *The Theory Underlying Concept Maps and How to Construct and Use Them*. Technical Report IHMC CmapTools 2006-01 Rev 01-2008. Florida Institute for Human and Machine Cognition, 2008 (cited on pages 21, 22, 91, 102).
- [152] Joseph D. Novak and Musonda Dismas. “A Twelve-Year Longitudinal Study of Science Concept Learning”. In: *American Educational Research Journal* 28.1 (1991), pages 117–153 (cited on pages 21, 102).

- [153] Stefanie Nowak and Stefan R  ger. “How Reliable are Annotations via Crowdsourcing”. In: *Proceedings of the International Conference on Multimedia Information Retrieval (MIR)*. 2010, pages 557–566 (cited on page 8).
- [154] Terence Odlin. *Language Transfer: Cross-Linguistic Influence in Language Learning*. Cambridge, New York, USA: Cambridge University Press, 1989 (cited on pages 28, 29).
- [155] David Oleson, Alexander Sorokin, Greg Laughlin, Vaughn Hester, John Le and Lukas Biewald. “Programmatic Gold: Targeted and Scalable Quality Assurance in Crowdsourcing”. In: *Proceedings of the AAAI Workshop on Human Computation (HCOMP)*. 2011, pages 43–48 (cited on page 51).
- [156] Abraham Othman and Tuomas Sandholm. “Decision Rules and Decision Markets”. In: *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*. 2010, pages 625–632 (cited on page 16).
- [157] Tim Paek, Yun-cheng Ju and Christopher Meek. “People Watcher: A Game for Eliciting Human-Transcribed Data for Automated Directory Assistance”. In: *Proceedings of the Annual Conference of the International Speech Communication Association (INTERSPEECH)*. 2007 (cited on page 17).
- [158] Aditya Parameswaran and Neoklis Polyzotis. “Answering Queries using Humans, Algorithms and Databases”. In: *Proceedings of the Conference on Innovative Data Systems Research (CIDR)*. Asilomar, CA, USA, 2011 (cited on page 12).
- [159] Manoj Parameswaran and AB Whinston. “Research Issues in Social Computing”. In: *Journal of the Association for Information Systems* 8.6 (2007), pages 336–350 (cited on page 9).
- [160] Manoj Parameswaran and Andrew B. Whinston. “Social Computing: An Overview”. In: *Communications of the Association for Information Systems* 19 (2007), pages 762–780 (cited on page 8).
- [161] David L Passmore, Evin Dogan Cebeci and Rose M Baker. “Potential of Idea Futures Markets in Educational Technology”. In: *Proceedings of the 5th International Conference on Information Technology-Based Higher Education and Training*. Istanbul, Turkey, 2004 (cited on page 15).
- [162] Lisa Pearl and Mark Steyvers. “Identifying Emotions, Intentions, and Attitudes in Text Using a Game with a Purpose”. In: *Proceedings of NAACL-HLT Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*. 2010, pages 71–79 (cited on page 17).
- [163] Massimo Poesio, Jon Chamberlain, Udo Kruschwitz, Livio Robaldo and Luca Ducceschi. “Phrase Detectives: Utilizing Collective Intelligence for Internet-Scale Language Resource Creation”. In: *ACM Transactions on Interactive Intelligent Systems* 3.1 (2013) (cited on page 17).

- [164] Drazen Prelec. “A Bayesian truth serum for subjective data.” In: *Science* 306.5695 (2004) (cited on page 14).
- [165] Liangfei Qiu, Huaxia Rui and Andrew B. Whinston. “Information Exchange in Prediction Markets: How Social Networks Promote Forecast Efficiency”. In: *Proceedings of the Hawaii International Conference on System Sciences (HICSS)*. 2013 (cited on page 15).
- [166] Alexander J Quinn and Benjamin B Bederson. “Human Computation: A Survey and Taxonomy of a Growing Field”. In: *Proceedings of the 29th ACM Conference on Human Factors in Computing Systems (CHI)*. 2011 (cited on pages 8–13).
- [167] M. Jordan Raddick, Georgia Bracey, Pamela L. Gay, Chris J. Lintott, Phil Murray, Kevin Schawinski, Alexander S. Szalay and Jan Vandenberg. “Galaxy Zoo: Exploring the Motivations of Citizen Science Volunteers”. In: *Astronomy Education Review* 9.1 (2010) (cited on page 19).
- [168] Walter Rafelsberger and Arno Scharl. “Games with a Purpose for Social Networking Platforms”. In: *Proceedings of the 20th ACM conference on Hypertext and Hypermedia (HT)*. Torino, Italy, 2009, pages 193–197 (cited on page 13).
- [169] Rick Raymer. “Gamification: Using Game Mechanics to Enhance eLearning”. In: *eLearn* (2011) (cited on page 21).
- [170] Francesco Ricci, Lior Rokach, Bracha Shapira and Paul B. Kantor. *Recommender Systems Handbook*. Boston, MA: Springer, 2011 (cited on page 49).
- [171] P. Clint Rogers, Stephen W. Liddle, Peter Chan, Aaron Doxey and Brady Isom. “Web 2.0 Learning Platform: Harnessing Collective Intelligence”. In: *Turkish Online Journal of Distance Education* 8.3 (2007), pages 16–33 (cited on page 21).
- [172] C. Romero and S. Ventura. “Educational Data Mining: A Survey from 1995 to 2005”. In: *Expert Systems with Applications* 33 (2007), pages 135–146 (cited on page 21).
- [173] Marta Sabou, Kalina Bontcheva and Arno Scharl. “Crowdsourcing Research Opportunities: Lessons from Natural Language Processing”. In: *Proceedings of the 12th International Conference on Knowledge Management and Knowledge Technologies (i-KNOW)*. 2012 (cited on page 17).
- [174] Yuko Sakurai, Tenda Okimoto, Masaaki Oka, Masato Shinoda and Makoto Yokoo. “Quality-Control Mechanism Utilizing Worker’s Confidence for Crowdsourced Tasks (Extended Abstract)”. In: *Proceedings of the 12th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*. Saint Paul, Minnesota, USA, 2013, pages 1347–1348 (cited on page 11).
- [175] Katie Salen and Eric Zimmerman. *Rules of Play: Game Design Fundamentals*. MIT Press, 2004 (cited on pages 55, 73).

- [176] Daniel Schall. “Crowdsourcing Task Marketplaces”. In: *Service-Oriented Crowdsourcing*. SpringerBriefs in Computer Science. New York, NY: Springer, 2012, pages 7–30 (cited on page 18).
- [177] Andrew I Schein, Alexandrin Popescul, Lyle H Ungar and David M Pennock. “Methods and Metrics for Cold-Start Recommendations”. In: *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2002 (cited on page 71).
- [178] Doug Schuler. “Social Computing”. In: *Communications of the ACM* 37.1 (1994), pages 28–29 (cited on page 8).
- [179] Emile Servan-Schreiber, Justin Wolfers, David M Pennock and Brian Galebach. “Prediction Markets: Does Money Matter?” In: *Electronic Markets* 14.3 (2004), pages 243–251 (cited on page 15).
- [180] David R Shanks. “Learning: From Association to Cognition”. In: *Annual Review of Psychology* 61 (2010), pages 273–301 (cited on pages 4, 99).
- [181] S S Shapiro and M B Wilk. “An Analysis of Variance Test for Normality”. In: *Biometrika* 52.3/4 (1965), pages 591–611 (cited on page 110).
- [182] George Siemens and Phil Long. “Penetrating the Fog: Analytics in Learning and Education”. In: *Educause Review* 46.5 (2011) (cited on pages 21, 112).
- [183] M Six Silberman, Joel Ross, Lilly Irani and Bill Tomlinson. “Sellers’ problems in human computation markets”. In: *Proceedings of the ACM SIGKDD Workshop on Human Computation (HCOMP)*. 2010, pages 18–21 (cited on page 18).
- [184] Jakub Šimko and Mária Bieliková. “Games with a Purpose: User Generated Valid Metadata for Personal Archives”. In: *Proceedings of the 6th International Workshop on Semantic Media Adaptation and Personalization*. 2011, pages 45–50 (cited on page 71).
- [185] Katharina Siorpaes and Martin Hepp. “Games with a Purpose for the Semantic Web”. In: *IEEE Intelligent Systems* 23.3 (2008), pages 50–60 (cited on pages 13, 71).
- [186] David Smiley and Eric Pugh. *Solr 1.4 Enterprise Search Server*. Packt Publishing, 2009 (cited on page 123).
- [187] Gregory Smith. *PostgreSQL 9.0 High Performance*. Packt Publishing, 2010 (cited on page 120).
- [188] John B. Smith. *Collective Intelligence in Computer-Based Collaboration*. Hillsdale, New Jersey, USA: Lawrence Erlbaum Associates, 1994 (cited on page 7).
- [189] Rion Snow, Brendan O’Connor, Daniel Jurafsky and Andrew Y Ng. “Cheap and Fast — But is it Good? Evaluating Non-Expert Annotations for Natural Language Tasks”. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2008 (cited on page 11).

- [190] Erik Snowberg, Justin Wolfers and Eric Zitzewitz. “Partisan Impacts on the Economy: Evidence from Prediction Markets and Close Elections”. In: *The Quarterly Journal of Economics* 122.2 (2007), pages 807–829 (cited on page 15).
- [191] Robert Speer, Catherine Havasi and Harshit Surana. “Using Verbosity: Common Sense Data from Games with a Purpose”. In: *Proceedings of the 23rd International Florida Artificial Intelligence Research Society Conference (FAIRS)*. 2010, pages 104–109 (cited on page 13).
- [192] Bartholomäus Steinmayr, Christoph Wieser, Fabian Kneissl and François Bry. “Karido: A GWAP for Telling Artworks Apart”. In: *Proceedings of the 16th International Conference on Computer Games (CGAMES)*. 2011, pages 193–200 (cited on pages 14, 25).
- [193] Stanley Smith Stevens. “On the Theory of Scales of Measurement”. In: *Science* 103.2684 (1946), pages 677–680 (cited on page 35).
- [194] Tarja Susi, Mikael Johannesson and Per Backlund. *Serious Games – An Overview*. Technical Report. University of Skövde, School of Humanities and Informatics, 2007 (cited on page 19).
- [195] Stefan Thaler, Elena Simperl, Katharina Siorpaes and Christian Hofer. *A Survey on Games for Knowledge Acquisition*. Technical Report. Semantic Technology Institute Innsbruck, 2011 (cited on page 14).
- [196] Ei Pa Pa Pe-Than, Dion Hoe-Lian Goh and Chei Sian Lee. “A Typology of Human Computation Games: An Analysis and a Review of Current Games”. In: *Behaviour & Information Technology* November 2013 (2013) (cited on page 14).
- [197] Deborah J Trumbull, Rick Bonney, Derek Bascom and Anna Cabral. “Thinking Scientifically during Participation in a Citizen-Science Project”. In: *Science Education* 84.2 (2000), pages 265–275 (cited on page 19).
- [198] Donna Vakharia and Matthew Lease. *Beyond AMT: An Analysis of Crowd Work Platforms*. arXiv:1310.1672v1. University of Texas at Austin, 2013. arXiv: 1310.1672v1 (cited on page 11).
- [199] Linda O. Valenty and Ofer Feldman, editors. *Political Leadership for the New Century: Personality and Behavior Among American Leaders*. Praeger Frederick, 2002 (cited on page 45).
- [200] Laura L. Veldkamp. “Information Markets and the Comovement of Asset Prices”. In: *Review of Economic Studies* 73.3 (2006), pages 823–845 (cited on page 15).
- [201] Aditya Vempaty, Lav R. Varshney and Pramod K. Varshney. “Reliable Classification by Unreliable Crowds”. In: *Proceedings of the International Conference on Acoustics, Speech, and Signal (ICASSP)*. 2013, pages 5558–5562 (cited on page 8).
- [202] Alessandro Vinciarelli, Maja Pantic and Hervé Bourlard. “Social Signal Processing: Survey of an Emerging Domain”. In: *Image and Vision Computing* 27.12 (2009), pages 1743–1759 (cited on page 9).



- [203] Luis von Ahn. “Games with a Purpose”. In: *Computer* 39.6 (2006), pages 92–94 (cited on page 13).
- [204] Luis von Ahn. “Human Computation”. PhD thesis. Carnegie Mellon University, 2007 (cited on page 12).
- [205] Luis von Ahn and Laura Dabbish. “Designing Games with a Purpose”. In: *Communications of the ACM* 51.8 (2008), pages 58–67 (cited on pages 13, 14, 71).
- [206] Luis von Ahn and Laura Dabbish. “Labeling Images with a Computer Game”. In: *Proceedings of the 22nd ACM Conference on Human Factors in Computing Systems (CHI)*. 2004, pages 319–326 (cited on page 13).
- [207] Maja Vukovic and Claudio Bartolini. “Towards a Research Agenda for Enterprise Crowdsourcing”. In: *Leveraging Applications of Formal Methods, Verification, and Validation*. Edited by Tiziana Margaria and Bernhard Steffen. Lecture Notes in Computer Science 6415. Springer, 2010, pages 425–434 (cited on page 10).
- [208] Bo Waggoner and Yiling Chen. “Information Elicitation Sans Verification”. In: *Proceedings of the 3rd Workshop on Social Computing and User Generated Content at the ACM Conference on Electronic Commerce*. 2013 (cited on pages 8, 14).
- [209] Fei-Yue Wang, Daniel Zeng, Kathleen M. Carley and Wenji Mao. “Social Computing: From Social Informatics to Social Intelligence”. In: *IEEE Intelligent Systems* 22.2 (2007), pages 79–83 (cited on page 9).
- [210] Wendy A. Warr. “Social Software: Fun and Games, or Business Tools?” In: *Journal of Information Science* 34.4 (2008), pages 591–604 (cited on page 10).
- [211] Ingmar Weber, Stephen Robertson and Milan Vojnovic. “Rethinking the ESP game”. In: *Proceedings of the 27th ACM Conference on Human Factors in Computing Systems (CHI)*. New York, New York, USA, 2009 (cited on page 13).
- [212] Albert Weichselbraun, Stefan Gindl and Arno Scharl. “Using Games with a Purpose and Bootstrapping to Create Domain-Specific Sentiment Lexicons”. In: *Proceedings of the 20th ACM International Conference on Information and Knowledge Management (CIKM)*. Glasgow, Scotland, UK, 2011, pages 1053–1060 (cited on page 14).
- [213] Elizabeth T. Welsh, Connie R. Wanberg, Kenneth G. Brown and Marcia J. Simmering. “E-learning: emerging uses, empirical results and future directions”. In: *International Journal of Training and Development* 7.4 (2003), pages 245–258 (cited on page 20).
- [214] Christoph Wieser, François Bry, Alexandre Bérard and Richard LaGrange. “ARTigo: Building an Artwork Search Engine With Games and Higher-Order Latent Semantic Analysis”. In: *Proceedings of Disco, Workshop on Human Computation and Machine Learning in Games at HCOMP*. 2013 (cited on page 73).

- [215] Jens Witkowski and David C Parkes. “A Robust Bayesian Truth Serum for Small Populations”. In: *Proceedings of the 13th ACM Conference on Electronic Commerce (EC)*. 2012 (cited on page 14).
- [216] Jens Witkowski and David C Parkes. “Learning the Prior in Minimal Peer Prediction”. In: *Proceedings of the 3rd Workshop on Social Computing and User Generated Content at the ACM Conference on Electronic Commerce*. 2013 (cited on page 14).
- [217] Jens Witkowski and David C Parkes. “Peer Prediction without a Common Prior”. In: *Proceedings of the 13th ACM Conference on Electronic Commerce (EC)*. 2012, pages 964–981 (cited on page 14).
- [218] Justin Wolfers and Eric Zitzewitz. *Five Open Questions About Prediction Markets*. IZA Discussion Paper, number 1975. 2006 (cited on page 16).
- [219] Justin Wolfers and Eric Zitzewitz. “Prediction Markets”. In: *Journal of Economic Perspectives* 18.2 (2004), pages 107–126 (cited on page 15).
- [220] Justin Wolfers and Eric Zitzewitz. *Prediction Markets in Theory and Practice*. IZA Discussion Paper, number 1991. 2006 (cited on page 15).
- [221] Anita Williams Woolley, Christopher F Chabris, Alex Pentland, Nada Hashmi and Thomas W Malone. “Evidence for a Collective Intelligence Factor in the Performance of Human Groups”. In: *Science* 330 (2010), pages 686–688 (cited on page 7).
- [222] Jiang Yang, Lada A Adamic and Mark S Ackerman. “Crowdsourcing and Knowledge Sharing: Strategic User Behavior on Taskcn”. In: *Proceedings of the 9th ACM Conference on Electronic Commerce (EC)*. 2008, pages 246–255 (cited on page 18).
- [223] Michael Juntao Yuan, Jacob Orshalick and Thomas Heute. *Seam Framework: Experience the Evolution of Java EE*. Prentice Hall, 2009 (cited on pages 115, 116, 118).
- [224] Man-Ching Yuen, Irwin King and Kwong-Sak Leung. “A Survey of Crowdsourcing Systems”. In: *Proceedings of the IEEE 3rd International Conference on Social Computing*. Boston, MA, USA, 2011, pages 766–773 (cited on page 11).
- [225] Rabih Zbib, Erika Malchiodi, Jacob Devlin, David Stallard, Spyros Matsoukas, Richard Schwartz, John Makhoul, Omar F Zaidan and Chris Callison-Burch. “Machine Translation of Arabic Dialects”. In: *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics*. 2012 (cited on page 17).
- [226] Daqing Zhang, Bin Guo and Zhiwen Yu. “The Emergence of Social and Community Intelligence”. In: *Computer* 24.7 (2011), pages 21–28 (cited on page 9).

- [227] Matthew Zook, Mark Graham, Taylor Shelton and Sean Gorman. “Volunteered Geographic Information and Crowdsourcing Disaster Relief: A Case Study of the Haitian Earthquake”. In: *World Medical & Health Policy* 2.2 (2010), pages 7–33 (cited on page [10](#)).



## INTERNET REFERENCES

- [228] *Accenti Urbani*. URL: <http://labs.play4science.org/accentiurbani> (visited on 01/02/2014) (cited on page 115).
- [229] *Adobe Flash Player*. URL: <https://www.adobe.com/software/flash/about/> (visited on 01/02/2014) (cited on page 45).
- [230] Christopher Allen. *Tracing the Evolution of Social Software*. 2004. URL: [http://www.lifewithalacrity.com/2004/10/tracing%5C\\_the%5C\\_evo.html](http://www.lifewithalacrity.com/2004/10/tracing%5C_the%5C_evo.html) (visited on 01/02/2014) (cited on page 10).
- [231] *Amazon Mechanical Turk*. URL: <https://www.mturk.com> (visited on 01/02/2014) (cited on page 11).
- [232] *Apache Lucene - Apache Solr*. URL: <https://lucene.apache.org/solr> (visited on 01/02/2014) (cited on page 123).
- [233] *ARTigo*. URL: <http://www.artigo.org> (visited on 01/02/2014) (cited on pages 14, 25, 73, 115).
- [234] Eric Bidelman. *Capturing Audio & Video in HTML5*. 2013. URL: <http://www.html5rocks.com/en/tutorials/getusermedia/intro/> (visited on 01/02/2014) (cited on page 45).
- [235] François Bry. *Nicht normierende Massmedien*. 2011. URL: <http://www2.pms.ifi.lmu.de/erlebt/?p=5241> (visited on 01/02/2014) (cited on page 28).
- [236] Tom Coates. *My Working Definition of Social Software...* 2003. URL: [http://plasticbag.org/archives/2003/05/my%5C\\_working%5C\\_definition%5C\\_of%5C\\_social%5C\\_software](http://plasticbag.org/archives/2003/05/my%5C_working%5C_definition%5C_of%5C_social%5C_software) (visited on 01/02/2014) (cited on page 10).
- [237] *CrowdFlower*. URL: <http://crowdfower.com> (visited on 01/02/2014) (cited on page 52).
- [238] *CrowdWorx*. URL: <http://www.crowdworx.com> (visited on 01/02/2014) (cited on page 17).
- [239] Gerard E. Dallal. *The Little Handbook of Statistical Practice*. 2012. URL: <http://www.jerrydallal.com/LHSP/LHSP.htm> (cited on page 110).
- [240] *Duolingo*. URL: <http://www.duolingo.com> (visited on 01/02/2014) (cited on page 18).
- [241] *Facebook*. URL: <https://www.facebook.com> (visited on 01/02/2014) (cited on page 46).
- [242] *Folding@home*. URL: <http://folding.stanford.edu> (visited on 01/02/2014) (cited on page 12).

- [243] *Galaxy Zoo*. URL: <http://www.galaxyzoo.org> (visited on 01/02/2014) (cited on page 19).
- [244] *Gamification Wiki*. URL: <http://gamification.org> (visited on 01/02/2014) (cited on page 20).
- [245] *GraphViz - Graph Visualization Software*. URL: <http://www.graphviz.org> (visited on 01/02/2014) (cited on page 110).
- [246] *Handbook of Collective Intelligence*. URL: [http://scripts.mit.edu/~cci/HCI/index.php?title=Main%5C\\_Page](http://scripts.mit.edu/~cci/HCI/index.php?title=Main%5C_Page) (visited on 01/02/2014) (cited on page 8).
- [247] *Hibernate. Everything data*. URL: <http://hibernate.org> (visited on 01/02/2014) (cited on pages 120, 121).
- [248] Jeff Howe. *Crowdsourcing*. URL: <http://crowdsourcing.typepad.com> (visited on 01/02/2014) (cited on page 10).
- [249] Jeff Howe. *The Rise of Crowdsourcing*. 2006. URL: <http://www.wired.com/wired/archive/14.06/crowds.html> (cited on pages 3, 10).
- [250] *IMDb - Movies, TV and Celebrities*. URL: <http://www.imdb.com> (visited on 01/02/2014) (cited on page 104).
- [251] *Intrade*. URL: <http://www.intrade.com> (visited on 01/02/2014) (cited on page 17).
- [252] Katharina Jakob. *www.metropolitalia.org - Giocate con noi?* 2012. URL: <http://italians.corriere.it/2012/10/07/www-metropolitalia-org-giocate-con-noi/> (visited on 01/02/2014) (cited on page 77).
- [253] *JBoss Application Server 7 - JBoss Community*. URL: <https://www.jboss.org/jbossas> (visited on 01/02/2014) (cited on page 118).
- [254] *JeuxDeMots*. URL: <http://www.jeuxdemots.org> (visited on 01/02/2014) (cited on page 17).
- [255] *List of Wikipedias*. URL: [https://meta.wikimedia.org/wiki/List%5C\\_of%5C\\_Wikipedias](https://meta.wikimedia.org/wiki/List%5C_of%5C_Wikipedias) (visited on 01/02/2014) (cited on page 28).
- [256] *Lumenogic*. URL: <http://www.lumenogic.com> (visited on 01/02/2014) (cited on page 17).
- [257] Maxmind. *GeoLite Free Downloadable Databases*. URL: <http://dev.maxmind.com/geoip/legacy/geolite/> (visited on 01/02/2014) (cited on page 53).
- [258] *Media Capture and Streams*. URL: <http://www.w3.org/TR/2013/WD-mediacapture-streams-20130516> (visited on 01/02/2014) (cited on page 45).
- [259] *MediaStream Recording*. URL: <http://www.w3.org/TR/2013/WD-mediastream-recording-20130205> (visited on 01/02/2014) (cited on page 45).
- [260] *Metropolitalia*. URL: <http://www.metropolitalia.org> (visited on 01/02/2014) (cited on pages 4, 25).

- [261] *metropolititalia.org* / *Facebook*. URL: <https://www.facebook.com/Metropolititalia> (visited on 01/02/2014) (cited on page 77).
- [262] *Phrase Detectives*. URL: <http://www.phrasedetectives.org> (visited on 01/02/2014) (cited on page 17).
- [263] *Piwik*. URL: <http://piwik.org> (visited on 01/02/2014) (cited on page 81).
- [264] *Play4Science*. URL: <http://www.play4science.org> (visited on 01/02/2014) (cited on pages 25, 73).
- [265] *PostgreSQL: The world's most advanced open source database*. URL: <http://www.postgresql.org> (visited on 01/02/2014) (cited on page 120).
- [266] *Seam Framework - JBoss Seam*. URL: <http://seamframework.org> (visited on 01/02/2014) (cited on pages 115, 116).
- [267] *SETI@home*. URL: <http://setiathome.ssl.berkeley.edu> (visited on 01/02/2014) (cited on page 12).
- [268] Bill Shannon. *Java Platform, Enterprise Edition (Java EE) Specification, v5*. Technical report. Sun Microsystems, 2006. URL: <https://jcp.org/aboutJava/communityprocess/final/jsr244/index.html> (cited on page 116).
- [269] Clay Shirky. *Social Software and the Politics of Groups*. 2003. URL: [https://web.archive.org/web/20120204035008/http://www.shirky.com/writings/group%5C\\_politics.html](https://web.archive.org/web/20120204035008/http://www.shirky.com/writings/group%5C_politics.html) (visited on 01/02/2014) (cited on page 10).
- [270] *Stack Overflow*. URL: <http://www.stackoverflow.com> (visited on 01/02/2014) (cited on pages 10, 18).
- [271] *TeX - LaTeX Stack Exchange*. URL: <http://tex.stackexchange.com> (visited on 01/02/2014) (cited on page 104).
- [272] *The Java Community Process(SM) Program - JSRs: Java Specification Requests - detail JSR#244*. URL: <https://jcp.org/en/jsr/detail?id=244> (visited on 01/02/2014) (cited on page 115).
- [273] *Twitter*. URL: <https://www.twitter.com> (visited on 01/02/2014) (cited on page 46).
- [274] *Wikimedia users - Strategic Planning*. URL: [http://strategy.wikimedia.org/wiki/Wikimedia%5C\\_users?oldid=44715](http://strategy.wikimedia.org/wiki/Wikimedia%5C_users?oldid=44715) (visited on 01/02/2014) (cited on page 81).
- [275] *Wikipedia*. URL: <http://en.wikipedia.org> (visited on 01/02/2014) (cited on page 10).
- [276] *Zooniverse - Real Science Online*. URL: <https://www.zooniverse.org> (visited on 01/02/2014) (cited on page 19).